Based on Latest Examination Pattern of
Lucknow University, Lucknow

PARICHAY®
CLASS NOTES

Including
Compulsory
Question
Based on
Unit Pattern

B.Sc. I Year
STATISTICS
(Probability Distributions and Numerical Analysis)
PAPER - II

- According to new pattern of Question papers short notes (Compulsory Question) and absorptive Questions based on unit pattern.

Dr. J. S. Manglik

New Edition

UNITECH
UNITECH HOUSE
Aminabad, Lucknow-226 018

₹ 160.00

CONTENT

Compulsory Questions

UNIT : 1 1–13

(1) Discuss Poisson Distribution as a limiting case of Binomial Distribution.

(2) Discuss the applicability of Rolle's theorem for the following functions :

(i) $f(x) = \sin x \ \forall \ [0, \pi]$.

(ii) $f(x) = \cos 1/x \ \forall x \in [-1, 1]$.

(iii) $f(x) = \tan x \ \forall x \in [0, \pi]$.

(3) If $f(x)$ be a function such that $f'(x) = 0$ for all x in (a, b) then prove that $f(x)$ reduces to a constant in (a, b).

(4) If two functions $f(x)$ and $\phi(x)$ have the same derivative for each x in (a, b) then prove that they differ by a constant in (a, b).

(5) What are the conditions for application of Binomial Distribution ?

(6) What are the characteristics of Binomial Distribution ?

(7) What are the conditions under which Poisson Distribution is used ?

(8) What are the characteristics of Poisson Distribution ?

(9) Delineate Binomial distribution and its important features.

(10) Explicate Poisson distribution and its properties.

(11) Explain and define the term Binomial Distribution. What are the physical conditions for Binomial distribution?

(12) Write a short note on Negative Binomial Distribution.

(13) Show that the mode of the Poisson distribution with parameter m lies between $(m - 1)$ and m. Can it

coincide with $(m-1)$ or m ? If so, when and which of these values will be the mode ?

(14) A Poisson distribution has a double mode at $x = 4$ and $x = 5$. Find the probability that x will have either of these values.

(15) If x is a Poisson variate such that $P(x = 1) = P(x = 2)$ then evaluate $P(x = 4)$

(16) If x and y are Poisson variates with means m_1 and m_2, prove that the probability that $(x - y)$ has the value r is the coefficient of t^r in exp $(m_1 t + m_2 t^{-1} - m_1 - m_2)$.

(17) Prove that the sum of two independent Gamma variates with parameters l and m is a gamma variate with parameter $l + m$.

UNIT : II

50–66

(1) What do you understand by chi-square distribution. What are its properties ?

(2) What are the uses of Chi-square tests ?

(3) What are the conditions applying the chi-square tests ?

(4) Explain Chi-square distribution of variance.

(5) Describe the chi-square test of significance and state the various uses to which it can be put. **Or**

What is χ^2 test ? Point out its uses.

(6) What do you mean by F-statistic ? Discuss the derivation of Snedecor's F-distribution. What are the constants of F-distribution?

(7) What do you mean by Random Variable ? Give the statement of some of the theorems on Random Variables.

(8) What do you mean by distribution function ? Prove that if F is the distribution function of one dimensional random variable X, then

$$f(-\infty) = \lim_{x \to -\infty} f(x) = 0$$

and $$f(\infty) = \lim_{x \to \infty} f(x) = 1.$$

(9) Write a short note on Bivariate Normal Distribution.

(10) What do you understand by student-distribution ? What are the various properties of t-distribution ?

(11) Discuss the Fisher's f-Test for testing the equality of two sample variance.

(12) Write Conditions for the application of χ^2 test.

(13) Numerical Question

(14) Numerical Question

(15) Write about the Student's t-distribution.

(16) Find student's t for the following variate in a sample of eight taking the mean of the universe to be zero.

$-4, -2, -2, 0, 2, 2, 3, 3.$

(17) Ten individuals are chosen at random from a normal population and their heights are found to be

63, 63, 66, 67, 68, 69, 70, 70, 71, 71 inches.

Test if the sample belongs to the population whose mean height is 66 inches. Given t for –9 d.f. at 5% level of significance is 2.262.

(18) Numerical Question

(19) Write M.G.F. of Chi-Square Distribugion

(20) Additive or Re-productive Property of Noncentral Chi-Square Distribution : If $Y_i, (i = 1, 2, \ldots, k)$, are independent non-central χ^2-variates with $n_i\, d.f.$ and non-centrality element λ_i then $\sum_{i=1}^{k} \lambda_i$.

UNIT : III

101–108

(1) Drive a relation Between E and Δ and E^{-1} and ∇

(2) $E\,\Delta \equiv \Delta\, E$ and $E\,\nabla \equiv \nabla\, E \equiv \Delta$

(3) Prove $D \equiv \frac{1}{h} \log(1 + \Delta)$

(4) $\delta\,[f(x)\,g(x)] = \mu\, f(x)\,\delta\, g(x) + \mu\, g(x)\,\delta\, f(x)$

(5) Write output the Calculus of Finite Difference and Operator

(6) $$u_x - \frac{1}{8}\Delta^2 u_{x-1} + \frac{1\cdot 3}{8\cdot 16}\Delta^4 u_{x-2} - \frac{1\cdot 3\cdot 5}{8\cdot 16\cdot 24}\Delta^3 u_{x-3} + \ldots.$$

$$= u_{x+1/2} - \frac{1}{2}\Delta u_{x+1/2} + \frac{1}{4}\Delta^2 u_{x+1/2} - \frac{1}{8}\Delta^3 u_{x+1/2} + \ldots$$

(7) Given$u_0 = 3, u_1 = 12, u_2 = 81, u_3 = 2000, u_4 = 100, u_5 = 8$ find $\Delta^5 u_0$.

(8) Find $f(6)$ given that $f(0) = -3, f(1) = 6, f(2) = 8, f(3) = 12$ the third difference being constant.

(9) Show that $\underset{yz}{\Delta^2} x^3 = x + y + z = f(x, y, z), f(x) = x$

(10) Numerical Question

(11) What do you known about Central Difference formula.

(12) Evaluate $\frac{\Delta^2}{E} \sin(x+h) + \frac{\Delta^2 \sin(x+h)}{E \sin(x+h)}$, h being the interval of differencing.

(13) If $u_x = ax^2 + bx + c$, then show that

$u_{2n} - {}^nC_1 2u_{2n-1} + {}^nC_2 2^2 u_{n-2} - \ldots + (-2)^n u_n = (-1)^n (c - 2an)$

(14) Given that $u_x = e^{ax+b}$, find $\Delta^n u_x$.

(15) Express a given polynomial in factorial notation.

(16) Give a brief Introduction of divided differences.

UNIT : IV

128–135

(1) Evaluate $\int_0^{10} \frac{dx}{1+x^2}$ By Weddle's rule.

(2) Numerical Question

(3) Numerical Question

(4) The equation $\sin x = 5x - 2$ can be put as $x = \sin^{-1}(5x - 2)$ and also as $x = \frac{1}{5}(\sin x + 2)$ suggesting two iterative process for its solution which of these if any would succeed and which would fail to give the root in the neighbourhood of $0 \cdot 5$.

(5) Starting with $x = \cdot 12$, solve $x = \cdot 21 \sin(\cdot 5 + x)$ by using iteration method.

(6) Obtain e^{-1}, where $e = 2 \cdot 71$ starting with $x_0 = 0 \cdot 3$.

(7) Give Table of Comparison Between different Methods.

(8) Explain the iteration method

(9) Numerical Question

(10) Calculate (upto 3 places of decmila) $\int_2^{10} \frac{dx}{1+x}$ by dividing the range into eight equal parts.

(11) Evaluate $\int_0^4 e^x\,dx$, by Simpson's rule, using the data $e = 2\cdot72, e^2 = 7\cdot39, e^3 = 20\cdot09, e^4 = 54\cdot60,$ and compare it with the actual value.

(12) Calculate $\int_0^{\pi/2} e^{\sin x}\,dx$ correct to four decimal places.

(13) Numerical Quiestion

(14) Find the value of $\log_e 2$ from $\int_0^1 \frac{x^2}{1+x^3}\,dx$, using Simpson's '$\frac{1}{3}$' rule, by dividing the range into four equal parts.
Find the error also.

(15) Evaluate $\int_0^1 e^{-x^2}\,dx$ By Simpson's rule.

(16) Find the Moment of Inertia of the area about the x-axis .

Long Answer Type Questions

UNIT : I

1. What do you understand by Binomial Distribution ? Calculate the mean and variance of Binomial Distribution. 13
2. Explain and define Poisson Distribution. Calculate the mean and variance of Poisson Distribution. 16
3. Discuss elaborately Normal distribution and its characteristics. 17
4. What is hypergeometric distribution ? Discuss means and variance of hypergeometric distribution. 20
5. Discuss Gamma distribution, Multinomial distribution, Exponential distribution and Beta distribution. 22
6. What do you understand by Normal Distribution ? Discuss normal distribution as a limiting form of Binomial Distribution. 27
7. If x is normally distributed with zero mean and unit variance, find the expectation and variance of σ^2 31
8. In a normal distribution whose mean is 2 and S.D. 3, find a value of the variate such that the probability of the interval from the mean to the value is 0.4115. Find another value such

that the probability for the interval from $x = 3 \cdot 5$ to the value is 0.2307. 32
9. In a normal distribution 31% of items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution. 33
10. For the Beta variate of the first kind with parameters m and n, evaluate Harmonic Mean. 33
11. Write about Log-normal Distribution. 34
12. Write about Standard Laplace (Double Exponential) Distribution. 35
13. Write about the Weibul Distribution. 37
14. What is Cauchy Distribution and Write Characteristics of it. 39
15. Define exponential distribution and obtain its mean and variance. 42
16. Obtain recurrence relation for finding moments of Poisson distribution. What is the additive property of independent Poisson variates ? Prove it. 42
17. Define negative binomial distribution Find its m.g.f. and hence find its mean and variance. 45
18. Discuss the main features of a normal distribution. 46
19. For a rectangular distribution :

$$f(x) = \frac{1}{2a}, -a \le x \le a$$

Show that : $\mu_{2r} = \frac{a^{2r}}{2r+1}$ 47
20. Define a Poisson distribution. Obtain the recurrence formula :

$$\mu_{r+1} = r\lambda\mu_{r-1} + \lambda \frac{d\mu_r}{d\lambda}$$

for this distribution and hence, obtain the first four moments. 48
21. Show that in a Poisson distribution with unit mean, mean deviation about mean is $(2/e)$ times the standard deviation. 49

UNIT : II

1. If χ_1^2 and χ_2^2 are two independent Chi-square variates with n_1 and n_2 degrees of freedom respectively, then show that $\left(\chi_1^2 + \chi_2^2\right)$ is also a Chi-square variate with $(n_1 + n_2)$ degrees of freedom. 66
2. Define F-statistic and give the properties of F-distribution. 66

3. Define χ^2-distribution. Find its moment generating function andhence show that its variance is two times the mean. 6

4. Explain in detail Student's t-distribution. 68

5. Explain in detail Chi-Square Distribution. 77

6. Discuss F-Distribution in Detail. 85

7. Explain the uses of t-distribution in testing the hypothesis.

Or

Discuss the procedure in which the 'student' distribution is generally used to test the significance of the various results obtained from small samples. 87

8. The joint p.d.f. $guv\,(u, v)$ of the transformed variables U and V is : $guv\,(u, v) = f_{XY}\,(x, y)\,|\,J\,|$, where $|\,J\,|$ is the modulus value of the Jacobian of transformation and $f\,(x, y)$ is expressed in terms of u and v. 90

9. If X and Y are independent continuous $r.v.$'s, then the $p.d.f.$ of $U = X + Y$ is given by :

$$h\,(u) = \int_{-\infty}^{\infty} f_X\,(v)\, f_Y\,(u - v)\, dv$$

90

10. Numerical Question 91

11. Distribution of the Product of two Random Variables : Let X and Y be two independent continuous $r.v.$'s. Then the $p.d.f.$ of their product $U = XY$, is given by :

$$h\,(u) = \int_{-\infty}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{dx}{|\,x\,|}.$$

93

12. Theorem : Distribution of the Quotient of Two Random Variables : If X and Y are independent continuous random variables, then the $p.d.f.$ of the quotient $Z = X/Y$, is given by :

$$g_1\,(z) = \int_{-\infty}^{\infty} g\,(vz, v)\,|\,v\,|\,dv = \int_{-\infty}^{\infty}$$

$$f_X\,(vz)\, f_Y\,(v)\,|\,v\,|\,dv$$

94

13. Let (X, Y) be a two-dimensional non-negative continuous $r.v.$, having the joint density :

$$f\,(x, y) = \begin{cases} 4xy\, e^{-(x^2 + y^2)}; & x \geq 0, y \geq 0 \\ 0, & elsewhere \end{cases}$$

Prove that the density function of $U = \sqrt{X^2 + Y^2}$ is :

$U = \sqrt{X_2 + Y^2}$ is :

$$h\,(u) = \begin{cases} 2u^3\, e^{-u^2}, & 0 \leq u < \infty \\ 0, & elsewhere \end{cases}$$

94

14. Given the joint density function of X and Y as :

$$f(x,y)=\begin{cases}1/2\,xe^{-y}\,; & 0<x<2, y>0\\ 0, & \text{elsewhere}\end{cases}$$

find the distribution of $X+Y$ 96

15. Theorem. If X_1 and X_2 are two independent χ^2-variates with n_1 and n_2 $d.f.$ respectively, then $\frac{X_1}{X_2}$ is $\alpha\ \beta_2\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate. 97

16. If X_1 and X_2 are independent χ^2-variates with n_1 and n_2 d.f. respectively, then $U=\frac{X_1}{X_1+X_2}$ and $V=X_1+X_2$ are independently distributed, U as $\beta_1\left(\frac{n_1}{2},\frac{n_2}{2}\right)$ variate and V as a χ^2-variate with $(n_1+n_2)\,d.f.$ 98

17. Write Conditions for the validity of χ^2-test. 99

18. Establish a Relation Between F and χ^2 distribution. 99

UNIT : III

1. Discuss the Gauss's Forward Formula. 108
2. Discuss the Gauss Backward Formula 110
3. Numerical Question 110
4. Numerical question 111
5. From the following
sin 45° = 0.7071, sin 50° = 0.7660
sin 55° = 0.8192, sin 60° = 0.8660
estimate sin 52° by using any method of interpolation. 112
6. Numerical Question 113
7. Explain the Divided Differences. 114
8. The divided differences are symmetrical in all their arguments, that is, the value of any difference is independent of the order of the arguments. 115
9. The nth divided difference can be expressed as the quotient of two determinants each of order $n+1$. 117
10. The nth divided difference can be expressed as the quotient of two determinants each of order $n+1$. 117
11. Numerical Questions 119
12. Establish Relation Between Divided Differences and Ordinary Differences. 121
13. Numerical Question 121

14. Show that $\underset{bcd}{\Delta^3}\left(\frac{1}{a}\right) = -\frac{1}{abcd}$. 122

15. Given $\log_{10} 654 = 2 \cdot 8156$, $\log_{10} 658 = 2 \cdot 8182$,
$\log_{10} 659 = 2 \cdot 8189$, $\log_{10} 661 = 2 \cdot 8202$;
find $\log_{10} 656$. 123

16. Numerical Questions 124

17. Four equidistant values u_{-1}, u_0, u_1 and u_2 being given, a value is interpolated by Lagrange's formula. Show that it may be written in the form

$$u_x = yu_0 + xu_1 + y\frac{(y^2-1)}{3\,!}\Delta^2 u_{-1} + \frac{x(x^2-1)}{3\,!}\Delta^2 u_0,$$

where $x + y = 1$. 124

18. The values of $f(x)$ are given at a, b and c. Show that the maximum or minimum is attained by

$$x = \frac{f(a)(b^2-c^2) + f(b)(c^2-a^2) + f(c)(a^2-b^2)}{2\{f(a)(b-c) + f(b)(c-a) + f(c)(a-b)\}}.$$ 125

19. Numerical Question 126

20. Using Lagnange's interpolation formula find a polynomial which passes the points (0, –12), (1, 0), (3, 6), (4, 12). 126

UNIT : IV

1. What is Newton's backward interpolation formula ? 135

2. Derive general quadrature formula for numerical integration. 136

3. By means of Lagrange's formula prove that.
$y_1 = y_3 - 0.3(y_5 - y_{-3}) + 0.2(y_{-3} - y_{-5})$ 137

4. Derive the Weddle rule of numerical integration. 138

5. Explain Simpson's 3/8 rule for numerical integration. 138

6. Numerical Question 139

7. Numerical Question 140

8. Find the root of the equation $2x = \cos x + 3$ correct to three decimal places. 141

9. Find the cube root of 10 by itertive method. 141

10. Find a real root of the equation

$$f(x) = x^3 + x^2 - 1 = 0$$

by using iteration method. 142

11. Discuss Method of Successive Approximation or Iteration. 143

12. Let $x = \alpha$ be a root of $f(x) = 0$, which is equivalent to $x = \phi(x)$ and I be an interval containing α. If $|\phi'(x)| < 1$, for

all x in I, then the sequence of approximations $x_0, x_1, \ldots, x_n$ will converge to the root α, provided that the initial approximation x_0 is chosen in I. 144

13. Find a real root of $\cos x = 3x - 1$, correct to three decimal places. 145

14. Find by the iteration method, the root near 3.8, of the equation $2x - \log_{10} x = 7$ correct to four decimal places. 146

15. Discuss the Newton-Raphson Method. 146

16. Find by Newton's method the real root of the equation $3x = \cos x + 1$ 146

17. Using Newton's iterative method, find the real root of $x \log_{10} x = 1 \cdot 2$ correct to five decimal places, 147

18. Calculate the total arc of the ellipse $4x^2 + y^2 = 4$. 148

19. Compute the value of the definite integral $\int_{0\cdot 2}^{1\cdot 4} (\sin x - \log_e x + e^x)\, dx$ by (i) Trapezoidal rule (ii) Simpson's 1/3 – rule (iii) Simpson's 3/8 – rule (iv) Weddle rule. After finding the true value of the integral, compare the error in the four cases. 149

20. Numerical Question 150

21. Prove Simpson's formula

$$\int_a^b f(x)\, dx = \frac{b-a}{6n} [f(x_0) + 4f(x_1) + 2f(x_2) + \ldots + f(x_{2n})]$$

where $x_0 = a, x_{2n} = b$

and use it to evaluate $\int_1^2 \frac{dx}{x}$ and give estimates of error for $n = 1$ and 2, given that $\log_e 2 = 0 \cdot 69315$. 151

UNIT 1

Your Syllabus

Univariate distributions : Binomial, Poisson, Hypergeometric, Geometric and Negative Binomial. Uniform (discrete & continuous), Normal, Exponential, Gamma, Beta distributions. Cauchy, Laplace, Pareto, Weibull, Log normal Distributions. Normal and Poisson distributions as limiting case of binomial distribution.

Compulsory Questions

(1) Discuss Poisson Distribution as a limiting case of Binomial Distribution.

Ans. Poisson Distribution (as a limiting case of Binomial Distribution). Poisson distribution was discovered by the French mathematician and physicist Simeon Densi Poisson (1781- 1840) who published it in 1837. Poisson distribution is a limiting case of the binomial distribution under the following conditions :

(i) n, the number of trials is indefinitely large, i.e., $n \to \infty$.

(ii) p, the constant probability of success for each trial is indefinitely small, *i.e.*, $p \to 0$.

(iii) $np = \lambda$, (say), is finite. Thus $p = \lambda/n, q = 1 - \lambda/n$, where λ is a positive real number.

The probability of x successes in a series of n independent trial is

$$b\,(x\,, n, p) = \binom{n}{x} p^x q^{n-x}\,; x = 0, 1, 2, \dots n \qquad \dots(1)$$

We want the limiting form of (1) under the above conditions. Hence

$$\lim_{n \to \infty} b\,(x\,; n, p) = \lim_{n \to \infty} \frac{n\,!}{x\,!\,(n-x)\,!} \left(\frac{\lambda}{n}\right)^x \left[1 - \frac{\lambda}{n}\right]^{n-x}$$

Using Stirling's approximation for $n\,!$ as $n \to \infty$ *viz.*,

$$\lim_{n \to \infty} n\,! = \sqrt{2\pi}\, e^{-n}\, n^{n + (1/2)},$$ we get

$$\lim_{n \to \infty} b\,(x\,; n, p) = \lim_{n \to \infty} \left[\frac{\sqrt{2\pi}\, e^{-n} \cdot n^{n + (1/2)}}{x\,!\,\sqrt{2\pi}\, e^{-(n-x)} \cdot (n-x)^{n - x + (1/2)}}\right]$$

$$\left(\frac{\lambda}{n}\right)^x\left[1-\frac{\lambda}{n}\right]^{n-x}$$

$$=\frac{\lambda^x}{e^x\cdot x\,!}\cdot\lim_{n\to\infty}\frac{n^{n-x+(1/2)}}{(n-x)^{n-x+(1/2)}}\cdot\left[1-\frac{\lambda}{n}\right]^{n-x}$$

$$=\frac{\lambda^x}{e^x x\,!}\cdot\lim_{n\to\infty}\frac{\left(1-\frac{\lambda}{n}\right)^{n-x}}{\left(1-\frac{x}{n}\right)^{n-x+(1/2)}}$$

$$=\frac{\lambda^x}{e^x x\,!}\;\frac{\lim\limits_{n\to\infty}\left[1-\frac{\lambda}{n}\right]^n\cdot\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-x}}{\lim\limits_{n\to\infty}\left[1-\frac{x}{n}\right]^n\lim\limits_{n\to\infty}\left[1-\frac{x}{n}\right]^{-x+(1/2)}}$$

But we know that

and
$$\left.\begin{aligned}&\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}\\&\lim_{n\to\infty}\left[1-\frac{\lambda}{n}\right]^{\alpha}=1,\ \alpha\text{ is not a function of }n\end{aligned}\right\}\qquad\text{...(2)}$$

Therefore

$$\lim_{n\to\infty}b\,(x\,;n,p)=\frac{\lambda^x}{e^x\cdot x\,!}\cdot\frac{e^{-\lambda}\cdot 1}{e^{-x}\cdot 1}$$

$$=\frac{e^{-\lambda}\cdot\lambda^x}{x\,!}\;;x=0,1,2,\ldots,\infty;\qquad\text{[Using (2)]}$$

which is required probability function of the Poisson distribution. 'λ' is known as the parameter of Poisson distribution. ●

➲ **(2) Discuss the applicability of Rolle's theorem for the following functions :**

(i) $f(x)=\sin x\ \forall\ [0,\pi]$.

(ii) $f(x)=\cos 1/x\ \forall\ x\in[-1,1]$.

(iii) $f(x)=\tan x\ \forall\ x\in[0,\pi]$.

Ans. (i) Since $\sin x$ is continuous and derivable for all values of x, it is continuous on $[0,\pi]$ and is derivable on $(0,\pi)$.

Also $f(0)=0=f(\pi)$. Thus all conditions of Rolle's theorem are satisfied.

Now $f'(x)=\cos x=0$ when $x=\pm\pi/2,\pm 3\pi/2$.

Since $\pi/2\in(0,\pi)$, Rolle's theorem is verified.

(ii) The function $\cos 1/x$ is discontinuous at $x=0$ (and hence it is not differentiable there). Therefore Rolle's theorem is not applicable, Note that $f(-1) = f(1)$.

(iii) Since $\pi/2 \in [0, \pi]$ and $\tan \pi/2 = \pm\infty$, $\tan x$ is discontinuous at $x = \pi/2$.

Hence the condition (i) of Rolle's theorem is not satisfied. Consequently Rolle's theorem is not applicable. ●

➲ (3) If $f(x)$ be a function such that $f'(x) = 0$ for all x in (a, b) then prove that $f(x)$ reduces to a constant in (a, b).

Ans. If x_1, x_2 be any two values of x in (a, b), then from Lagrange's Mean value Theorem, we have

$$f(x_1) - f(x_2) = (x_1 - x_2) f'(c), a \le c \le b$$

$$= 0, \qquad \because f'(x) = 0 \text{ for all } x \text{ in } (a, b)$$

Hence $f(x_1) = f(x_2)$ ...(1)

Now as x_1, x_2 are any two values of x in (a, b), so we conclude from (1) above that $f(x)$ has the same value for each x in (a, b) and so is a constant in (a, b). ●

➲ (4) If two functions $f(x)$ and $\phi(x)$ have the same derivative for each x in (a, b) then prove that they differ by a constant in (a, b).

Ans. Given that $f'(x) = \phi'(x)$, for each x in (a, b).

So $f'(x) - \phi'(x) = 0$ *i.e.,* the derivative of $[f(x) - \phi(x)]$ is zero for each x in (a, b).

$\therefore f(x) - \phi(x)$ reduces to a constant for each x in (a, b)

i.e., $f(x)$ and $\phi(x)$ differ by a constant in (a, b). ●

➲ (5) What are the conditions for application of Binomial Distribution ?

Ans. 1. The variable should be discrete, *i.e.,* the values of x could be 1, 2, 3, 4 or 5 etc. and never 1·5, 2·1 or 3·41 etc.

2. A dichotomy exists. In other words the happening of events must be of two alternatives. It must be either a success or a failure.

3. The number of trials n should be finite and small.

4. The trials or events must be independent. The happening of one event must not affect the happening of other events. In other words, Statistical Independence must exist.

5. The trials or events must be repeated under identical conditions. ●

➲ (6) What are the characteristics of Binomial Distribution ?

Ans. 1. It is a discrete distribution which gives the theoretical probabilities.

2. It depends on the parameters p or q, the probability of success or failure and n (the number of trials). The parameter n is always a positive integer.

3. The distribution will be symmetrical if $p = q$. It is skew-symmetric or asymmetric if $p \neq q$ although with n tending to be large it is approximately so.

4. The statistics of, the Binomial distribution are mean = np; Variance = npq; and standard deviation = $\sqrt{npq}$.

5. The mode of the Binomial distribution is equal to the value of X which has the largest frequency.

6. It can be represented graphically, taking the X-axis to represent the number of successes and Y-axis to represent the probabilities or frequencies. Its graph will be vertical lines, with spaces in between them. Drawing a smooth curve by free hand is inadmissible as the variable X is a discrete one.

7. The shape and location of a Binomial distribution changes as p changes for a given n or n changes for a given p.

8. The Binomial coefficients are given by the Pascal's Triangle. ●

➲ (7) What are the conditions under which Poisson Distribution is used ?

Ans. 1. The random variable X should be discrete.

2. A dichotomy exists, *i.e.*, the happening of the events must be of two alternatives such as success and failure; occurrence and non-occurrence etc.

3. It is applicable in those cases where the number of trials n is very large and the probability of success p is very small but the mean $np = m$ is finite.

4. p should be very small (close to zero). If $p \to 0$, then the distribution is J-shaped and unimodal.

5. Statistical independence is assumed. In other words, it is applicable to those cases where the happening of one event does not affect the happening of the other events.

●

➲ (8) What are the characteristics of Poisson Distribution ?

Ans. 1. Poisson distribution is a discrete distribution. It gives theoretical probabilities and theoretical frequencies of a discrete variable.

2. It depends mainly on the value of the mean m.

3. This distribution is positively skewed to the left. With the increase in the value of the mean m, the distribution shifts to the right and the skewness diminishes.

4. Its arithmetic mean in relative distribution is p and in absolute distribution is np.

5. If n is large and p is small, this distribution gives a close approximation to Binomial distribution. Since the arithmetic mean of Poisson is same as that of Binomial, so the Poisson distribution can be used instead of Binomial if n or p is not known.

6. Poisson distribution has only one parameter, *viz.*, m, the arithmetic mean. Thus the entire distribution can be determined once the arithmetic mean is known.

7. Poisson distribution is based on the following assumptions :

(i) Statistical independence is assumed, *i.e.*, the occurrence or non-occurrence of an event does not influence the other events.

(ii) The probability of happening of more than one event in a very small interval is negligible.

(iii) The probability of success for a small space or a short interval of time is proportional to the space or length of time interval as the case may be.

●

➲ **(9) Delineate Binomial distribution and its important features.**

Ans. Binomial distribution was invented by James Bernoulli which was posthumously published in 1713. Let n (finite) Bernoulli trials be conducted with probability 'p' of a success and 'q' of a failure. The probability of x successes out of n Bernoulli trials is given by

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

where $x = 0, 1, 2, \dots, n$

$0 \le p \le 1$ and $p + q = 1$

A dichotomous r.v., X having the probability mass function $f(x) = \binom{n}{x} p^x q^{n-x}$ is said to follow *Binomial distribution* and is usually denoted as $b(n, p)$ or $b(x; n, p)$. The probability function $\binom{n}{x} p^x q^{n-x}$ is the $(x+1)^{\text{th}}$ term in the binomial expansion of $(q+p)^n$.

Important Features

(i) If $n = 1$, the binomial distribution reduces to Bernoulli distribution. Hence, seldom Bernoulli distribution is known as *point binomial.*

(ii) Binomial distribution has two parameters n and p.

(iii) Mean of binomial distribution is np.

(iv) Variance of binomial distribution is npq.

(v) The moment generating function of binomial distribution is $(q + pe^t)^n$.

(vi) Characteristic function of $b(n, p)$ is

$$(q + pe^{it})^n.$$

(vii) The first moments with their relation to cumulants are :

$$\mu = \kappa_1 = np, \ \mu_2 = \kappa_2 = npq,$$

$$\mu_3 = \kappa_3 = npq\,(v - p)$$

$$\mu_4 = \kappa_4 + 3\kappa_2^2 = npq\,[1 + 3pq\,(n - 2)].$$

(viii) The measure of skewness, $\beta_1 = \frac{(1-2p)^2}{npq}$ or $\gamma_1 = \frac{1-2p}{\sqrt{npq}}$. The value of β_1 indicates that binomial distribution is positively skewed if $p < \frac{1}{2}$ and symmetrical if $p = \frac{1}{2}$.

(ix) $\beta_2 = 3 + \frac{1-6pq}{npq}$ or $\gamma_2 = \frac{1-6pq}{npq}$. The value of measure of kurtosis 'β_2' reveals that as the number of trials $n \rightarrow \infty$, the distribution tends to mesokurtic.

(x) Recurrence formula for binomial distribution is,

$$p(x = x+1) = p(x = x)\,\frac{n-x}{x+1}\cdot\frac{p}{q}$$

This formula is very helpful in calculating the probabilities of the consecutive successes.

(xi) If $X \sim b(n_1, p)$ and $Y \sim p(n_2, p)$ $X + Y \sim b(n_1 + n_2, p)$. This property is known as *additive or reproductive* property of binomial distribution.

(xii) If $Y \sim b(y; n, p)$, the variable $X = Y/n$, the ratio of number of successes to the total number of trials n, is called a *relative* or *Pseudo* binomial variate.

●

➲ (10) Explicate Poisson distribution and its properties.

Ans. Poisson distribution was discovered by a French Mathematician-cum-Physicist, Simeon Denis Poisson in 1937. He derived it as a limiting case of binomial distribution. If a dichotomous variable X is such that the constant probability p of success for each trial is very small and the number of trials n is indefinitely large and $np = \mu$ is finite, the probability of x successes is given by the probability mass function,

$$P(X = x) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots$$
$$= 0 \text{ otherwise}$$

It is denoted by $P(x; \mu)$.

Some of the example of Poisson variate are :

(a) No. of deaths in a city due to suicides.
(b) No. of defective items in a box of 100 items.
(c) No. of plane accidents per week.

Main properties of Poisson distribution are :

(i) Poisson distribution has only one parameter 'μ'.

(ii) Mean of Poisson distribution (variate) is μ.

(iii) Variance of Poisson distribution (variate) is also μ. This is the only distribution of which the mean and variance are equal.

(iv) Moment generating function of Poisson distribution is $e^{\mu(e^t - 1)}$.

(v) Characteristic function of Poisson distribution is $e^{\mu(e^{it}-1)}$.

(vi) First three moments of Poisson distribution (variate) are equal i.e., $\mu_1 = \mu_2 = \mu_3 = \mu$ and $\mu_4 = \mu + 3\mu^2$.

(vii) All the cumulants are equal i.e., $\kappa_1 = \kappa_2 = \kappa_3 = \kappa_4 = \mu$.

(viii) The measure of skewness, $\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{\mu^2}{\mu^3} = \frac{1}{\mu}$ or $\gamma_1 = \frac{1}{\sqrt{\mu}}$.

(ix) Measure of kurtosis, $\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{1}{\mu} + 3$ or $\gamma_2 = \frac{1}{\mu}$. These measures show that Poisson distribution is positively skewed and leptokurtic.

(x) For a Poisson variate X with mean μ, the recurrence relationship between $P(X = x)$ and $P(X = x + 1)$ is,

$$P(X = x + 1) = P(X = x) \cdot \frac{\mu}{x + 1}.$$

(xi) If n is large ($n \to \infty$) and p small in a binomial distribution such that np remains constant say, μ, the binomial distribution tends to Poisson distribution $\frac{e^{-\mu}\mu^x}{x!}$.

(xii) Independent Poisson variates posses additive (reproductive) property. This property ensures that the sum of independent oisson variates is also a Poisson variate. If $X_1, X_2, ..., X_n$ are $P(x_1 : \mu_1)$, $P(x_2; \mu_2)$, ... $P(x_n; \mu_n)$ respectively, then $(X_1 + X_2 + ... + X_n)$ is $P(\Sigma x_i; \mu_1 + \mu_2 + ... + \mu_n)$.

(xiii) If X and Y are two Poisson variates with means μ and λ respectively, the conditional distribution of X given $(X + Y)$ is binomial.

(xiv) If X_1 and X_2 are two independent Poisson variates distributed as $P(x_1; \mu_1)$ and $P(x_2; \mu_2)$ respectively, $(X_1 - X_2)$ is not a Poisson variate. This property of Poisson variates is known as the *subtractive* property.

●

➲ (11) Explain and define the term Binomial Distribution. What are the physical conditions for Binomial distribution?

Ans. Binomial Distribution. Binomial distribution was discovered by James Bernoulli (1654-1705) in the year 1700. Let a random experiment be performed repeatedly and let the occurrence of an event in a trial be called a success and its non-occurrence a failure. Consider a set of n independent Bernoullian trial (n being finite), in which the probability 'p' of success in any trial is constant for each trial. Then $q = 1 - p$, is the probability of failure in any trial.

The probability of x successes and consequently $(n - x)$ failures in n independent trials, in a specified order (say) *SSFSFFFS...FSF* (where

S represents success and F failure) is given by the compound probability theorem by the expression :

$$P(SSFSFFFS..FSF) = P(S)P(S)P(F)P(S)P(F)P(F)P(F)P(S) \times \ldots \times P(F)P(S)P(F)$$

$$= p.p.q.p.q.q.q.p\ldots q.p.q$$

$$= \underbrace{p.p.\ldots.p}_{x \text{ factors}} \quad \underbrace{q.q.q.\ldots.q}_{\{(n-x)\text{factors}\}} = p^x q^{n-x}$$

But x successes in n trials can occur in $\binom{n}{x}$ ways and the probability for each of these ways is $p^x q^{n-x}$. Hence the probability of x successes in n trials in any order whatsoever is given by the addition theorem of probability by the expression :

$$\binom{n}{x} p^x q^{n-x}$$

The probability distribution of the number of successes, so obtained is called the *Binomial probability distribution*, for the obvious reason that the probabilities of 0, 1, 2, ..., n successes, *viz.*,

q^n, $\binom{n}{1} q^{n-1} p$, $\binom{n}{2} q^{n-2} p^2, \ldots, p^n$, are the successive terms of the binomial expansion $(q+p)^n$.

Definition. *A random variable X is said to follow binomial distribution if it assumes only non-negative values and its probability mass function is given by*

$$P(X=x) = p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} \ ; x = 0, 1, 2, \ldots, n \ ; q = 1-p \\ 0, \text{otherwise} \end{cases} \quad \ldots(1)$$

The two independent constants n and p in the distribution are known as the *parameters* of the distribution. 'n' is also, sometimes, known as the degree of the binomial distribution.

Binomial distribution is a discrete distribution as, X can take only the integral values, *viz.*, 0, 1, 2, ..., n. Any variable which follows binomial distribution is known as *binomial variate*.

The probability $p(x)$ in (1) is also sometimes denoted by $b(x, n, p)$.

Notes : 1. This assignment of probabilities is permissible because

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (q+p)^n = 1.$$

2. Let us suppose that n trials constitute an experiment. Then if this experiment is repeated N times, the *frequency function* of the binomial distrbution is given by

$$f(x) = Np(x) = N\binom{n}{x} p^x q^{n-x} \ ; x = 0, 1, \ldots, n$$

and the expected frequencies of 0, 1, 2,..., n successes are the successive terms of the binomial expansion, $N(q+p)^n, q+p=1$.

Physical conditions for Binomial Distribution. We get the binomial distribution under the following experimental conditions.

(i) Each trial results in two mutually disjoint outcomes, termed as success and failure.

(ii) The number of trials 'n' is finite.

(iii) The trials are independent of each other.

(iv) The probability of success 'p' is constant for each trial.

The problems relating to tossing of a coin or throwing of dice or drawing cards from a pack of cards with replacement lead to binomial probability distribution.

Example. *Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.*

Solution. p = Probability of getting a head $= \frac{1}{2}$

q = Probability of not getting a head $= \frac{1}{2}$.

The probability of getting $x$heads in a random throw of 10 coins is $p(x) = \binom{10}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{10-x} = \binom{10}{x}\left(\frac{1}{2}\right)^{10}$;$x = 0, 1, 2, \ldots, 10$

∴ Probability of getting at least seven heads is given by

$$P(X \geq 7) = p(7) + p(8) + p(9) + p(10)$$

$$= \left(\frac{1}{2}\right)^{10}\left\{\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right\}$$

$$= \frac{120 + 45 + 10 + 1}{1024} = \frac{176}{1024}.$$

●

⊃ (12) Write a short note on Negative Binomial Distribution.

Ans. The equality of the mean and variance is an important characteristic of the Poisson distribution, whereas for the binomial distribution the mean is always greater than the variance. Occasionally, however, observable phenomena give rise to empirical discrete distributions which show a variance larger than the mean.

Some of the commonest examples of such behaviour are the frequency distributions of plant density obtained by quadrant sampling when the clustering of plants makes the simple Poisson model inapplicable. It has been shown by different investigators that in such cases the negative binomial distribution provides an excellent model because this distribution has a variance larger than the mean. Bacterial clustering (or contagion), *e.g.*, deaths of insects, number of insect bites leads to the negative binomial distribution and the distribution also arises in inverse sampling from a binomial population or as a weighted average of Poisson distribution. The negative binomial distribution can be derived drom empirical considerations in many ways.

Suppose we have a succession of n Bernoulli trails. We assume that (i) the trials independent (ii) the probability of success 'p' in a trial remains constant from trial to trial.

Let $f(x; r, p)$ denote the probability that there are x failures preceeding the rth success in $x + r$ trials.

Now, the last trial must be a success, whose probability is p. In the remaining $(x + r - 1)$ trials we must have $(r - 1)$ successes whose probability is given by

$$\binom{x+r-1}{r-1} p^{r-1} q^x.$$

Therefore by compound probability theorem, $f(x; r, p)$ is given by the product of these two probabilities, *i.e.*,

$$\binom{x+r-1}{r-1} p^{r-1} q^x \cdot p = \binom{x+r-1}{r-1} p^r q^x.$$

Definition. *A random variable X is said to follow a negative binomial distribution if its probability mass function is given by*

$$p(x) = p(X = x) = \binom{x+r-1}{r-1} p^r q^x; x = 0, 1, 2, \ldots$$
$$= 0, \text{ otherwise} \qquad \ldots(1)$$

$$\text{Also } \binom{x+r-1}{r-1} = \binom{x+r-1}{x} \qquad \left[\because \binom{n}{r} = \binom{n}{n-r}\right]$$

$$= \frac{(x+r-1)(x+r-2)\ldots(r+1)r}{x!}$$

$$= \frac{(-1)^x(-r)(-r-1)\ldots(-r-x+2)(-r-x+1)}{x!}$$

$$= (-1)^x \binom{-r}{x}$$

$$\therefore \quad p(x) = \begin{cases} \binom{-r}{x} p^r (-q)^x; x = 0, 1, 2, \ldots \\ 0, \textit{otherwise} \end{cases} \qquad \ldots(2)$$

which is the $(x + 1)^{th}$ term in the expansion of $p^r (1 - q)^{-r}$, a binomial expansion with a negative index. Hence the distribution is known as negative binomial distribution. Also

$$\sum_{x=0}^{\infty} p(x) = p^r \sum_{x=0}^{\infty} \binom{-r}{x} (-q)^x = p^r \times (1-q)^{-r} = 1$$

Therefore $p(x)$ represents the probability function and the discrete variable which follows this probability function is called the negative binomial variate.

If $p = \dfrac{1}{Q}$ and $q = \dfrac{P}{Q}$ so that $Q - P = 1$, $(\because p + q = 1)$

then $p(x) = \begin{cases} \left\{\dbinom{-r}{x} Q^{-r} \left(-\dfrac{P}{Q}\right)\right\}^{x} & ; x = 0, 1, 2, \ldots s \\ 0, \textit{otherwise} \end{cases}$ is the general term in the negative binomial expansion $(Q - P)^{-r}$. ●

➲ **(13) Show that the mode of the Poisson distribution with parameter *m* lies between (*m* – 1) and *m*. Can it coincide with (*m* – 1) or *m* ? If so, when and which of these values will be the mode ?**

Ans. We know that in Poisson distribution with parameter m, the probability of r successes

$$= (e^{-m} m^{r})/r!, \text{ where } r = 0, 1, 2, \ldots\ldots$$

Now we know that the mode of the above distribution is that value of r for which the above probability is greater than or equal to the term that precedes it and the term that follows it

$$\therefore \quad \frac{e^{-m} m^{r}}{r!} \geq \frac{e^{-m} m^{r-1}}{(r-1)!} \quad ..(i)$$

and

$$\frac{e^{-m} m^{r}}{r!} \geq \frac{e^{-m} m^{r+1}}{(r+1)!} \quad ..(ii)$$

From (i) we get $m \geq r$ or $r \leq m$ ..(iii)

From (ii) we get $(r + 1) \geq m$ or $m - 1 \leq r$ ..(iv)

$\therefore$ From (iii) and (iv) we get $(m - 1) \leq r \leq m$.

$\therefore$ If m is not an integer, then the mode r lies between $(m - 1)$ and m.

But if m is an integer, then either $(m - 1)$ or m is the mode, hence the mode can coincide with $(m - 1)$ or m if m is an integer. ●

➲ **(14) A Poisson distribution has a double mode at $x = 4$ and $x = 5$. Find the probability that x will have either of these values.**

Ans. Given that $P(x = 4) = P(x = 5)$ (Note) ..(i)

or

$$\frac{e^{-m} m^{4}}{(4)!} = \frac{e^{-m} m^{5}}{(5)!},$$

$$\because \quad P(x) = \frac{e^{-m} m^{x}}{(x)!},$$

for Poisson distribution

or

$$1 = \frac{m}{5}$$

or $$m = 5.$$

Now $$e^{-m} = e^{-5} = k0.006738, \text{ from tables,}$$

∴ $$P(x = 4) = (e^{-m} m^4)/(4)!$$

$$= \frac{(0.006738)(5)^4}{24} = \frac{4.211}{24} = 0.175 \qquad ...(iii)$$

∴ The required probability

$$= P(x = 4) + P(x = 5)$$

$$= 2P(x = 5), \text{ from (i)}$$

$$= 2(0.175), \text{ from (ii)}$$

$$= 0.350.$$

●

➲ **(15) If x is a Poisson variate such that $P(x = 1) = P(x = 2)$ then evaluate $P(x = 4)$**

Ans. Given that $P(x = 1) = P(x = 2)$

or $$\frac{e^{-m} m^1}{(1)!} = \frac{e^{-m} m^2}{(2)!},$$

∵ $$P(x) = \frac{e^{-m} m^x}{(x)!},$$

for Poisson distribution

or $$1 = \frac{1}{2} m$$

or $$m = 2.$$

∴ $$e^{-m} = e^{-2} \frac{1}{e^2} = \frac{1}{(2.7)^2} = \frac{100}{729} = 0.137$$

∴ $$P(x = 4) = \frac{e^{-m} m^4}{(4)!} = \frac{(0.137) \times 2^4}{4 \times 3 \times 2 \times 1}$$

$$= \frac{0.274}{3} = 0.0913$$

●

➲ **(16) If x and y are Poisson variates with means m_1 and m_2, prove that the probability that $(x - y)$ has the value r is the coefficient of t^r in $\exp(m_1 t + m_2 t^{-1} - m_1 - m_2)$.**

Ans. Here we are to find out the probability of $(x - y)$ taking the value r.

Now if $X - y = r$, then we can suppose that x takes the value $r + s$ and y the values s, where s can take the values 0, 1, 2,............

∴ $$P(x - y = r) = P(x = r + s, y = s) \qquad \text{(Note)}$$

$$= \sum_{s=0}^{\infty} \frac{e^{-m}{}_{1}m_1^{r+e}}{(r+s)!} \cdot \frac{e^{-m}{}_{2}m_2^{3}}{(s)!},$$

by the theorem of compound probability

$$= e^{-(m_1+m_2)} \sum_{s=0}^{\infty} \frac{m_1^{r+3} m_2^{8}}{(r+s)!\,(S)!}$$

Also $$e^{m_1 t + m_2 t^{-1}} = e^{m_1 t} \cdot e^{m_2 t^{-1}}$$

$$= \left[1 + m_1 t + \frac{(m_1 t)^2}{2!} + \frac{(m_1 t)^3}{3!} + \ldots.. + \frac{(m_1 t)^{r+s}}{(r+s)!} + \ldots\ldots\right]$$

$$\times \left[1 + m_2 t^{-1} + \frac{(m_2 t^{-1})^2}{2!} + \ldots + \frac{(m_2 t^{-1})^3}{s!} + \ldots.\right]$$

$\therefore$ Coefficient of t^r in $e^{m_1 t + m_2 t^{-1}} \sum_{s=0}^{\infty} \frac{m_1^{r+s}}{(r+s)!} \cdot \frac{m_2^{3}}{s!}$ **(Note)**

Hence from (i), we get

$$P(x - y = r) = e^{-m_1 - m_2} \times \text{coefficient of } t^r \text{ in } e^{m_1 t + m_2 t^{-1}}$$

$$= \text{coefficient of } t^r \text{ in } e^{m_1 t + m_2 t^{-1} - m_1 - m_2}$$

Hence proved

●

(17) Prove that the sum of two independent Gamma variates with parameters l and m is a gamma variate with parameter $l + m$.

Ans. Let x and y be two independent Gamma variates with parameters l and m respectively.

Also we know that m.g.f., of the sum of a number of independent variates is the product of their m.g.f.'s.

$\therefore$ m.g.f. of the sum of the above two independent Gamma variates is the product of their m.g.f.

$\therefore$ m.g.f. of their sum = product of $(1-t)^{-1}$ and $(1-t)^{-m}$

$$= (1-t)^{-1} \times (1-t)^{-m} = (1-t)^{-l-m}$$

which is the m.g.f. of a gamma variate with parameter $l + m$.

Hence Proved.

●

Long Answer Type Questions

Q. 1. What do you understand by Binomial Distribution ? Calculate the mean and variance of Binomial Distribution.

Ans. Binomial distribution is a discrete probability distribution which is obtained when the probability p of the happening of an event is same in all the trials, and there are only two events in each trial. For example, the probability of getting a head, when a coin is tossed a number of times, must remain same in each toss, *i.e.*, $p = \frac{1}{2}$.

Let an experiment consisting of n trials be performed and let the occurrence of an event in any trial be called a success and its non-occurrence a failure. Let p be the probability of success and q be the probability of the failure in a single trial, where $q = 1 - p$, so that $p + q = 1$.

Let us assume that trials are independent and the probability of success is same in each trial. Let us claim that we have n trials, then the probability of happening of an event r times and failing $(n - r)$ times in any specified order is $p^r q^{n-r}$ (by the theorem on multiplication of probability). But the total number of ways in which the event can happen r times exactly in n trials is $C(n, r)$. These $C(n, r)$ ways are equally likely, mutually exclusive and exhaustive.

Therefore, the probability of r successes and $(n - r)$ failures in n trials in any order, whatsoever is $= C(n, r)\, p^r q^{n-r}$.

It can also be expressed in the form

$$P(X = r) = P(r) = C(n, r)\, p^r q^{n-r};\ r = 0, 1, 2, 3, \ldots.\ n.$$

where $P(x = r)$ or $P(r)$ is the probability distribution a random variable X of the number of r successes. Giving different values to r, *i.e.*, putting $r = 0, 1, 2, 3, \ldots, n$, we get the corresponding probabilities ${}^nC_0 q^n, {}^nC_1 q^{n-1} p, {}^nC_2 q^{n-2} p^2, {}^nC_3 q^{n-3} p^3, \ldots, pn$, which are the different terms in the Binomial expansion of $(q + p)^n$.

As a result of it, the distribution $P(r) = C(n, r)\, p^r q^{n-r}$ is called Binomial Probability distribution. The two independent constants, *viz.*, n and p in the distribution are called the parameters of the distribution.

Again if the experiment (each consisting of n trials) be repeated N times, the frequency function of the Binomial distribution is given by

$$f(r) = N \times P(r) = N \times C(n, r)\, p^r q^{(n-r)}$$

The expected frequencies of 0, 1, 2, 3, ..., n successes in the above set of experiment are the successive terms in the Binomial expansion of $N(q + p)^n$; where $p + q = 1$, which is also called the Binomial frequency distribution.

Mean and Variance of Binomial Distribution

The probability distribution of Binomial distribution for r successes in n trials is given by

$$P(r) = {}^nC_r q^{n-1} p^r.$$

$$\text{Mean} = \mu = \sum_{i=0}^{i=n} x_i p_i = 0 . {}^nC_0 q^n + 1 . {}^nC_1 q^{n-1} p + 2 . {}^nC_2 q^{n-2} p^2 + \ldots. + n . {}^nC_n p^n$$

$$= nq^{n-1}p + 2 . \frac{n(n-1)}{2!} q^{n-2} p^2 + \dots + n . 1 p^n$$

$$\left(\because {}^nC_r = \frac{n!}{(n-r)! \times r!} \right)$$

$$= np\,(q^{n-1} + (n-1)\,q^{n-2} p + \dots + p^{n-1})$$

$= np\,(q + p)\,n - 1$ **(By Binomial Expansion)**

$= np.\ [\because (q + p) = 1]$

Hence mean $= np$.

Also variance $= \sigma^2 = \sum_{r=0}^{n} (x_r^2 p_r) - \mu^2$, where $p_r = P(r)$.

Here $x_r = r, r = 0, 1, 2, \dots, n$; so that

$$x_r^2 = r^2 = r + r(r-1), r = 0, 1, 2, \dots, n.\ (\because \mu = np)$$

$$\sigma^2 = \sum_{r=0}^{n} r^2 P(r) - n^2 p^2 = \sum_{r=0}^{n} [r(r-1) + r]\,{}^nC_r\, q^{n-r} p^r - n^2 p^2$$

$$[\because P(r) = {}^nC_r\, p^r q^{n-r}]$$

$$= \sum_{r=0}^{n} r(r-1) \frac{n!}{(n-r)! \times Xr!} p^r q^{n-r} + \sum_{r=0}^{n} r\,{}^nC_r\, q^{n-r} p^r - n^2 p^2.$$

$$= \sum_{r=0}^{n} r(r-1) \frac{n!}{(n-r)! \times r!} p^r q^{n-r} + \sum_{r=0}^{n} r\,{}^nC_r\, q^{n-r} p^r - n^2 p^2$$

$$= \sum_{r=0}^{n} \frac{n!}{(n-r)! \times (r-2)!} q^{n-r} p^r + np - n^2p^2$$

$$= n(n-1)p^2 \sum_{r=0}^{n} \frac{(n-2)!}{(n-r)! \times (r-2)!} q^{n-r} p^{n-r} + np - n^2p^2$$

$$= n(n-1)p^2 \Sigma\, {}^{n-2}C_{r-2}\, q^{(n-2)-(r-2)} p^{r-2} + np - n^2 p^2$$

$$= n(n-1)p^2 (q + p)^{n-2} + np - n^2 p^2$$

$$= n(n-1)p^2 + np - n^2 p^2 = (np - p + 1 - np)\,np$$

$$= np(1 - p) = npq.\ (\because 1 - p = q)$$

Hence Variance $(\sigma^2) = npq$.

Q. 2. Explain and define Poisson Distribution. Calculate the mean and variance of Poisson Distribution.

Ans. The Poisson distribution was discovered by a French Mathematician **Simen Denis Poisson in 1837. It is a discrete distribution** and is very widely used.

Poisson distribution is a limiting form of the Binomial distribution in which *n*, the number of trials, becomes very large and *p*, the probability of the success of the event is very small such that *np* is a finite quantity and not necessarily large. In the Binomial distribution, the probability of x successes is ${}^nC_x p^x q^{n-x}$. We require the limiting value of the expression $f(x) = {}^nC_x p^x q^{n-x}$, where $p = \frac{m}{n}$, where n ultimately tends to ∞ and m is a constant.

The probability distribution of a random variable *X* is said to have a Poisson Distribution if it takes only non-negative values and if its distribution is given by

$$P(X = x) = \frac{m^x e^{-m}}{x!}, x = 0, 1, 2, \ldots..,$$

where m is the parameter and

$$e = 2{\cdot}7183 = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!} + \ldots.$$

Hence the probability of 0, 1, 2, 3, ... x successes are given by

$$e^{-m}, \frac{e^{-m}\, m}{1!}, \frac{e^{-m}\, m^2}{2!}, \ldots., \frac{e^{-m}\, m^x}{x!}, \text{ respectively.}$$

Also the formula of Poisson Distribution for computing the theoretical or relative frequencies of a random variable x is given by

$$\text{Frequency} = N\left(\frac{e^{-m}\, m^x}{x!}\right), \text{ where } N \text{ is number of trials, } m = \text{mean.}$$

Derivation of Mean and Variance of Poisson Distribution

$$\text{Poisson Distribution is } P(r) = \frac{e^{-m} . m^r}{r!}, r = 0, 1, 2, \ldots.$$

$$\text{Mean } (\mu) = \sum_{i=0}^{i=n} x_i p_i = p_1 x_1 + p_2 x_2 + p_3 x_3 + \ldots + p_n x_n.$$

$$= oe^{-m} + |\, 1e^{-m} + 2\frac{m^2 e^{-m}}{2!} + 3\frac{m^3 e^{-m}}{3!} + 4\frac{m^4 e^{-m}}{4!} + \ldots.$$

$$= me^{-m} + m^2 e^{-m} + \frac{m^3 e^{-m}}{2 \times 1} + \frac{m^4 e^{-m}}{3 . 2 . 1} + \ldots.$$

$$= me^{-m}\left(1 + m + \frac{m^2}{2 . 1} + ...\right) = me^{-m}\, e^m$$

$$\left(\because e^m = 1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + ...\right)$$

$$= me^{-m+m} = me^0 = m$$

Also variance : $\sigma^2 = (\Sigma x_i^2 p_i) - \mu^2$

Now $\displaystyle\sum_{i=0}^{i=n} p_i x_i^2 = 0 . e^{-m} + 1 . me^{-m} + 2^2 . \frac{m^2 e^{-m}}{2!}$

$$+ 3^2 . \frac{m^3 e^{-m}}{3!} + 4^2 . \frac{m^4 e^{-m}}{4!} + ...$$

$$= me^{-m} + 2m^2 e^{-m} + \frac{3m^3 e^{-m}}{2 \times 1} + \frac{4m^4 e^{-m}}{3 \times 2 \times 1} + ...$$

$$= me^{-m}\left[1 + 2m + \frac{3m^2}{2 \times 1} + 4\frac{m^3}{3 \times 2 \times 1} + ...\right]$$

$$= me^{-m}\left[\left(1 + m + \frac{m^2}{2 \times 1} + \frac{m^3}{3 \times 2 \times 1} + ...\right) + \left(m + \frac{2m^2}{2} + \frac{3m^3}{3 \times 2 \times 1} + ...\right)\right]$$

$$= me^{-m}\left[e^m + m\left(1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + ...\right)\right]$$

$$= n^{e-m}(e^m + me^m) = m\,(m + 1)\, e^{-m} e^m = m\,(m + 1)$$

∴ Variance

$$= \Sigma x_i^2 p_i - \mu^2 = m\,(m + 1) - m^2 = m^2 + m - m^2 = m.$$

Standard deviation = $\sqrt{\text{Variance}} = \sqrt{m}$.

Q. 3. Discuss elaborately Normal distribution and its characteristics.

Ans. Normal distribution was first discovered by De-Moivre in 1733 and was also known to Laplace in 1774. Later it was derived by Kark Friedrich Gauss in 1809 and used it for the study of errors in astronomy. Any how, the credit of normal distribution has been given to Gauss and is often called Gaussion distribution. Normal distribution is the maximally used probability distribution in the theory of statistics.

A random variable X is said to follow a normal distribution with mean μ and variance σ^2, if its probability density function is

$$f_X(x;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

for $\quad -\infty < x < \infty$

$$-\infty < \mu < \infty \text{ and } \sigma > 0$$

The variate x is said to be distributed normally with mean μ and variance σ^2 and is denoted as $X \sim N(\mu, \sigma^2)$.

If $\mu = 0$ and $\sigma = 1$, then

$$f_X(x)=\frac{1}{\sqrt{2\pi}}e^{-1/2x^2}$$

Here, X is said standardised normal variate and is denoted as $X \sim N(0, 1)$. Also, p.d.f. $f_X(x)$ is called standardised normal distribution.

If $X \sim N(\mu, \sigma^2)$ and we make a transformation, $Z=\frac{X-\mu}{\sigma}$, the distribution of Z is a standardised normal distribution as $Z \sim N(0, 1)$ irrespective of the values of μ and σ^2 in the distribution of X. Z is also called *standard normal deviate*. We can write, $f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$.

Also, $\int_{-\infty}^{\infty} f(z)\,dz = 1$. The area under the standard normal curve between the ordinates at $z = 0$ and $z_1 = \phi(z)$

$$\phi(z)=\int_0^{z_1} f(z)\,dz.$$

Also $\quad \phi(z) = \phi(-z)$.

Characteristics of the normal distribution

(i) The normal distribution curve is bell shaped and is symmetrical about the line $x = \mu$.

(ii) The mode of the normal curve lies at the point $x = \mu$.

(iii) The area under the normal curve within its rage $-\infty$ to ∞ is always unity *i.e.,*

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,dx = 1.$$

(iv) On either side of the line $x = \mu$, the frequency decreases more rapidly within the range $(\mu \pm \sigma)$ and get slower and slower as we depart farther from the point μ on either side. The area under the normal curve beyond the distance $\pm 3\sigma$ is only $0 \cdot 27$ percent which is very small. As a matter of fact, the area under the normal curve within the range $\mu \pm \sigma$ is $0 \cdot 6826$, in the range $\mu \pm 2\sigma$ is $0 \cdot 9544$ and within the range $\mu \pm 3\sigma$ is $0 \cdot 9973$. The fact that $99 \cdot 73$ percent items are covered within the range

$\mu \pm 3\sigma$ of a normal distribution. It gives rise to the large number theory. Hence, for a sample of size $n \geq 30$, the distribution is taken as normal. The area under standard normal curve, bounded by the line $Z = \pm 1{\cdot}96$ is $0{\cdot}95$, *i.e.*, 95 percent. Since the normal curve is symmetrical, 2·5% area lies on the left tail and 2·5% on the right tail beyond the points $z = -1{\cdot}96$ and $z = 1{\cdot}96$, respectively.

(v) Mean = Median = Mode.

(vi) The normal curve is unimodal.

(vii) All odd order moments of the normal distribution are zero.

(viii) The first raw moment, *i.e.*, mean $= \mu$ and $\mu_2 = \sigma^2, \mu_3 = 0, \mu_4 = 3\sigma^4 = 3\mu_2^2$.

(ix) Measure of skewness, $\beta_1 = \dfrac{\mu_3^2}{\mu_2^2} = 0$ or $\gamma_1 = \sqrt{\beta_1} = 0$.

(x) Measure of Kurtosis, $\beta_2 = \dfrac{\mu_4}{\mu_2^2} = 3$ or $\gamma_2 = \beta_2 - 3 = 0$.

The value of β_1 and β_2 clearly reveal that the normal curve is symmetrical and mesokurtic.

(xi) Quartile deviation, Q.D. $= \dfrac{Q_3 - Q_1}{2} = \dfrac{2}{3}\sigma$.

(xii) The mean deviation about mean, $\text{M.D}_{\mu} =$

$$\int_{-\infty}^{\infty} |x - \mu| \, e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = \sqrt{\frac{2}{\pi}}\,\sigma = \frac{4}{5}\sigma.$$

(xiii) The maximum probability say, Max $[p(x)]$ of the normal curve occurs at the point $x = \mu$ whereas,

$$\text{Max}\,[p(x)] = \frac{1}{\sigma\sqrt{2\pi}}$$

As σ increases, $p(x)$ decreases and the curve becomes more and more flat and vice-versa.

(xiv) The moment generating function of the normal distribution, $N(x; \mu, \sigma^2)$ is

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = e^{\mu t + \frac{1}{2\sigma^2 t^2}}.$$

(xv) Characteristic function of the normal distribution, $N(x; \mu, \sigma^2)$ is

$$\phi_X(t) = \int_{-\infty}^{\infty} e^{itx} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx = e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

(xvi) The points of inflexion of the normal curve are given by $x = \mu \pm \sigma$ and at this point

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}.$$

(xvii) If $X_1, X_2, \ldots, X_n$ are n independent random variables distributed normally with mean $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ respectively, the sum $(X_1 + X_2 + \ldots\ldots + X_n)$ is distributed with mean $(\mu_1 + \mu_2 + \ldots. + \mu_n)$ and $(\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2)$.

(xviii) If X and Y are two normal variates such that $X \sim N(0, 1)$ and $Y \sim N(0, 1)$, the variate X/Y follow Cauchy distribution.

Q. 4. What is hypergeometric distribution ? Discuss means and variance of hypergeometric distribution.

Ans.

The Hypergeometric Distribution

Let n balls be drawn one at a time, *without replacement,* from an urn containing a white and b black balls. Then the probability that x of there n balls so drawn are white and remaining $(n - x)$ halls are black is given by

$$f(x) = \frac{{}^aC_x\, {}^bC_{n-x}}{{}^{a+b}C_n}, x = 0, 1, 2, \ldots, n \qquad \ldots(i)$$

$$\text{where } \Sigma f(x) = \sum_{x=0}^{n} \left[\frac{{}^aC_x\, {}^bC_{n-x}}{{}^{a+b}C_n}\right] = \frac{\sum\limits_{x=0}^{n} {}^aC_x\, {}^bC_{n-x}}{{}^{a+b}C_n} = \frac{{}^{a+b}C_n}{{}^{a+b}C_n} = 1 \quad \text{(Note)}$$

which can be proved by equating the coefficients of x^n on both sides in the product $(1 + x)^a (1 + x)^b = (1 + x)^{a+b}$ **(Note)**

$$\text{or} \qquad ({}^aC_0 + {}^aC_1 x + {}^aC_2 x^2 + \ldots\ldots + {}^aC_n x^n + \ldots)$$

$$({}^bC_0 + {}^bC_1 x + \ldots\, {}^bC_n x^n + \ldots$$

$$= 1 + {}^{a+b}C_1 x + {}^{a+b}C_2 x^2 + \ldots + {}^{a+b}C_n x^n + \ldots$$

Equating the coefficients of x^n on both sides we get

$${}^aC_0 + {}^bC_n + {}^aC_1\, {}^bC_{n-1} + {}^aC_2\, C_{n-2} + \ldots + {}^aC_n\, .\, {}^bC_0 = {}^{a+b}C_n$$

or $\sum\limits_{x=0}^{n} {}^aC_x\, {}^aC_{n-x} = {}^{a+b}C_n$, which gives the above result.

Hence $f(x)$ represents a probabilty density function.

The distribution with (i) as the probability law is called.

Mean of the Hypergeometric Distribution :

$$\text{Mean} = \mu_1' = \sum_{x=0}^{n} x f(x) = \sum_{x=0}^{n} x \frac{{}^aC_x\, {}^bC_{n-x}}{{}^{a+b}C_n}, \text{ from (i)}$$

$$= \frac{\sum\limits_{x=1}^{n} a\, {}^{a-1}C_{x-1}\, {}^bC_{n-x}}{{}^{a+b}C_n} = \frac{a}{{}^{a+b}C_n} \sum_{x=1}^{n} {}^{a-1}C_{x-1}\, {}^bC_{n-x}$$

$$= \frac{a}{{}^{a+b}C_n} \sum_{y=0}^{n-1} {}^{a-1}C_y \, {}^{b}C_{n-1-y}, \text{ where } x - 1 = y$$

$$= \frac{a}{{}^{a+b}C_n} \cdot {}^{a-1+b}C_{n-1}, \text{ as above}$$

$$= \frac{a\, n!\,(a+b-n)!}{(a+b)!} \; \frac{(a+b-1)!}{(n-1)!\,(a+b-1-n+1)!}$$

$$\because \; {}^{m}C_k = \frac{m!}{k!\,(m-k)!}$$

or mean $= \dfrac{an}{a+b}$. ...(i)

Variance of the Hypergeometric Distribution :

$$\mu_2' = \sum_{x=0}^{n} x^2 f(x) = \sum_{x=0}^{n} [x(x-1) + x] f(x) \qquad \textbf{(Note)}$$

or
$$\mu_2' = \sum_{x=0}^{n} x(x-1) f(x) + \sum_{x=0}^{n} x f(x)) \qquad \text{...(iii)}$$

Now $\displaystyle\sum_{x=0}^{n} x(x-1) f(x) = \frac{\sum_{x=0}^{n} x(x-1)\, {}^{a}C_x \, {}^{b}C_{n-x}}{{}^{a+b}C_n}$, from (i)

$$= \frac{a(a-1)}{{}^{a+b}C_n} \sum_{x=2}^{n} {}^{a-2}C_{x-2} \, {}^{b}C_{n-x}, \text{ as above}$$

$$= \frac{a(a-1)}{{}^{a+b}C_n} \sum_{y=0}^{n-2} {}^{a-2}C_y \, {}^{b}C_{n-y-2}, \text{ where } x - 2 = y$$

$$= \frac{a(a-1)}{{}^{a+b}C_n} \, {}^{a-2+b}C_{n-2}, \text{ as above}$$

$$= a(a-1) \frac{n!\,(a+b-n)!}{(a+b)!} \times \frac{(a+b-2)!}{(n-2)!\,(a+b-2-n+2)!}$$

$$= \frac{a(a-1)\,n(n-1)}{(a+b)(a+b-1)}$$

$\therefore$ From (iii), we get $\mu_2' = \dfrac{a(a-1)\,n(n-1)}{(a+b)(a+b-1)} + \dfrac{an}{a+b}$, ...(v)

substituting values from (ii), (iv)

$\therefore \mu_2 = \mu_2' - \mu_1'^{\,2}$

$$= \frac{a(a-1)\,n(n-1)}{(a+b)(a+b-1)} + \frac{an}{a+b} - \left(\frac{an}{a+b}\right)^2, \text{ from (ii), (v)}$$

$$= \frac{a(a-1)\,n(n-1)(a+b) + an(a+b)(a+b-1) - a^2n^2(a+b-1)}{(a+b)^2(a+b-1)}$$

$$= \frac{an\{(a-1)(n-1)(a+b) + (a+b)(a+b-1) - an(a+b-1)\}}{(a+b)^2(a+b-1)}$$

$$= \frac{an\{-a(a+b) - n(a+b) + (a+b)^2 + an\}}{(a+b)^2(a+b-1)}, \text{ on simplifying}$$

$$= \frac{an\{-bn + b^2 + ab\}}{(a+b)^2(a+b-1)} = \frac{abn(a+b-n)}{(a+b)^2(a+b-1)}$$

or varaince $= \mu_2 = \dfrac{abn(a+b-n)}{(a+b)^2(a+b-1)}$. ...(vi)

Q. 5. Discuss Gamma distribution, Multinomial distribution, Exponential distribution and Beta distribution.

Ans.

Gamma Distribution

We know that the gamma function Γn is defined for all real $n > 0$

as
$$\Gamma n = \int_0^\infty x^{n-1} e^{-x}\, dx. \qquad ...(i)$$

Gamma variate assumes all values from 0 to ∞ and is distributed with probability differential

$$dP = \frac{1}{\Gamma n} x^{n-1} e^{-x}\, dx, 0 \le x \le \infty.$$

Replacing x by az, the probability differential is given by

$$dP = \frac{1}{\Gamma n} (az)^{n-1} e^{-az} (a\, dz)$$

$$= \frac{1}{\Gamma n} a^n z^{n-1} e^{-az}\, dz = f(z)\, dz, 0 \le z \le \infty. \qquad ...(ii)$$

Again putting x for z, another form of the gamma distribution is

$$f(x)\, dx = \frac{1}{\Gamma n} a^n x^{n-1} e^{-ax}\, dx, 0 \le x \le \infty. \qquad ...(iii)$$

Note. For $n = 1$, the gamma distribution reduces to the exponential distribution having the parameter a. Note

m.g.f. for gamma distribution

$$f(x) = \frac{1}{\Gamma n} x^{n-1} e^{-x} \qquad ...(iv)$$

$$M_0(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x)\, dx$$

$$= \int_0^\infty e^{tx} \frac{1}{\Gamma n} x^{n-1} e^{-x}\, dx, \text{ from (iv)}$$

$$= \frac{1}{\Gamma n} \int_0^\infty x^{n-1} e^{-(-t+1)x}\, dx$$

$$= \frac{1}{\Gamma n} \int_0^\infty \left(\frac{z}{1-t}\right)^{n-1} e^{-z} \frac{dz}{1-t}, \text{ putting } x(1-t) = z,$$

$$dx = [1/(1-t)]\, dz \quad = \frac{1}{(1-t)^n \Gamma} \int_0^\infty z^{n-1} e^{-z}\, dz$$

$$= \frac{1}{(1-t)^n \Gamma} . \Gamma n, \text{ by definition} \qquad \text{[see result (i)]}$$

$$= (1-t)^{-n}$$

μ_r' for gamma distribution given by (iv) :

$$\mu_r' = E(X^r) = \int_0^\infty x^r f(x)\, dx$$

$$= \int_0^\infty x^r \frac{1}{\Gamma n} x^{n-1} e^{-x}\, dx, \text{ from (iv)}$$

$$= \frac{1}{\Gamma n} \int_0^\infty x^{n+r-1} e^{-x}\, dx = \frac{1}{\Gamma n} \Gamma(n+r), \text{ by (i)} \quad \text{...(v)}$$

Mean, Variance and moments about mean :

Putting $r = 1, 2, 3, \ldots$ in succession in (v) we get

$$\therefore \quad \mu_1' = \frac{1}{\Gamma n} \Gamma(n+1) = \frac{n!}{(n-1)!} = n$$

or $\quad$ **mean** $= \mu_1' = n$ ...(vi)

$$\therefore \quad \mu_2' = \frac{1}{\Gamma n} \Gamma(n+2) = \frac{(n+1)!}{(n-1)!} = (n+1)n \quad \text{...(vii)}$$

$$\therefore \quad \text{variance} = \mu_2 = \mu_2' - \mu_1'^2 = (n+1)n - n^2 = n \quad \text{...(viii)}$$

$$\mu_3' = \frac{1}{\Gamma n} \Gamma(n+3) = \frac{(n+2)!}{(n-1)!} = (n+2)(n+1)n \quad \text{...(ix)}$$

$$\mu_4' = \frac{1}{\Gamma n} \Gamma(n+4) = \frac{(n+3)!}{(n-1)!} = (n+3)(n+2)(n+1)n. \quad \text{...(x)}$$

From these we get

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3$$

$$= (n+2)(n+1)n - 3(n+1)n.n + 2n^3, \text{ from (vi), (vii), (ix)}$$

$$= n[(n+2)(n+1) - 3(n+1)n + 2n^2]$$

$$= n[n^2 + 3n + 2 - 3n^2 - 3n + 2n^2] = 2n \quad \text{...(xi)}$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^{2} - 3\mu_2'^{4}$$

$$= (n+3)(n+2)(n+1)n - 4(n+2)(n+1)n \,.\, n + 6(n+1)n\,.\,n^2 - 3n^4$$

$$= n(n+1)(n+2)\{(n+3) - 4n\} + 3n^3\{2(n+1) - n\}$$

$$= n(n+1)(n+2)(3-3n) + 3n^3(n+2)$$

$$= n(n+2)[(n+1)(3-3n) + 3n^2]$$

$$= n(n+2)[3n + 3 - 3n^2 - 3n + 3n^2] = 3n(n+2) \qquad \text{...(xii)}$$

Person's coefficients β_1 and β_2 :

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(2n)^2}{(n)^3}, \text{ from (xi) and (viii)}$$

or $\beta_1 = 4/n.$ ...(xiii)

$$\text{And } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3n(n+2)}{(n)^2}, \text{ from (viii), (xii)}$$

$$= 3 + \frac{6}{n}. \qquad \text{...(xiv)}$$

Standard Deviation for Gamma Distribution given by (iv) :

Standard Deviation $\sigma = \sqrt{(\mu_2)} = \sqrt{n}$, from (viii)

Exponential Distribution

For this distribution the probability density function is given by $f(x) = ce^{-x/\theta}, \theta > 0$ and the range of the exponential variate is $0 \le x \le \infty$.

Since it is a probability distribution, so

$$1 = \int_0^\infty f(x)\,dx = \int_0^\infty ce^{-x/\theta}\,dx = \left[ce^{-x/\theta}(-\theta)\right]_0^x$$

or $1 = 0 - (-c\theta) = c\theta$, which gives $c = 1/\theta$.

$\therefore$ The probability density function of an exponential variate reduces to

$$f(x) = (1/\theta)\,e^{-x/\theta}, \theta > 0, 0 \le x \le \infty.$$

m.g.f. for Exponential Distribution :

$$M_0(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x)\,dx = \int_0^\infty e^{tx}\left(\frac{1}{\theta}e^{-x/\theta}\right)dx$$

$$= \frac{1}{\theta}\int_0^\infty e^{-x[(1/\theta)-t]}\,dx = \frac{1}{\theta}\left[\frac{e^{-x[(1/\theta)-t]}}{-\{(1/\theta)-t\}}\right]_0^\infty$$

$$= \frac{1}{\theta}\left[0 + \frac{1}{(1/\theta)-t}\right] = \frac{1}{\theta} \times \frac{\theta}{1-\theta t} = (1-\theta t)^{-1} \qquad \text{...(i)}$$

μ_r' for Exponential Distribution :

From (i) $M_0(t) = 1 + \theta t + \theta^2 t^2 + \theta^3 t^3 + \ldots + \theta^r t^r + \ldots.,$

on expanding $(1 - \theta t)^{-1}$

$\therefore \quad \mu_r' =$ coefficient of $t^r/r!$ in $M_0(t)$

or $\quad \mu_r' = \theta^r\,(r!)$. ...(ii)

Substituting $r = 1, 2, 3, 4, \ldots$ in succession in (ii), we get

$$\mu_1' = \theta^1\,(1!) = \theta,\ \mu_2' = \theta^2\,(2!) = 2\theta^2,$$

$$\mu_3' = \theta^3\,(3!) = 6\theta^3,\ \mu_4' = \theta^4\,(4!) = 24\theta^4.$$

Mean and Variance of Exponential Distribution :

Mean $= \mu_1' = \theta$ ...(iii)

Also $\quad \mu_2 = \mu_2' = 2\theta^2 - (\theta)^2 = \theta^2$

$\therefore$ Variance $= \sqrt{\mu_2} = \theta$ or $\sigma = \theta$.

For this reason generally the probability density function of an exponential distribution is given by

$$f(x) = (1/\sigma)\,e^{-x/\sigma},\ \sigma > 0,\ 0 \le X \le \infty.$$

And for this form of distribution we can find as above

m.g.f. $= M_0(t) = (1 - \sigma t)^{-1}$ ...(iv)

$\mu_r' = \sigma^r\,(r!)$ ...(v)

Mean $= \mu_1' = \sigma$ ...(vi)

$$\mu_2' = 2\sigma^2,\ \mu_3' = 6\sigma^3,\ \mu_4' = 24\sigma^4.$$

Moments about mean of Exponential Distribution :

$\mu_1 = 0$, by definition

$$\mu_2 = \mu_2' - \mu_1'^2 = 2\sigma^2 - \sigma^2 = \sigma^2$$

$\therefore$ Variance $= \sqrt{(\mu_2)} = \sigma$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = 6\sigma^3 - 3\,(2\sigma^2)\,\sigma + 2\,(\sigma)^3 = 2\sigma^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

$$= 24\sigma^4 - 4\,(6\sigma^3)\,(\sigma) + 6\,(2\sigma^2)\,(\sigma)^2 - 3\,(\sigma)^4 = 9\sigma^4$$

Also Pearson's coefficients are

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(2\sigma^3)^2}{(\sigma^2)^3} = 4$$

and

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{9\sigma^4}{(\sigma^2)^2} = 9.$$

Beta Distribution of the first kind

The probability density function of the Beta variate of the first kind is given by

$$f(x) = \frac{1}{\beta(m, n)}\,x^{n-1}(1-x)^{m-1},\ 0 \le x \le 1,\ m, n > 0 \qquad \text{...(i)}$$

where $\beta(m, n) = \dfrac{\Gamma m \, \Gamma n}{\Gamma(m+n)} = \displaystyle\int_0^1 x^{m-1}(1-x)^{n-1}\,dx$ ...(ii)

The density given by (i) is also briefly referred as $\beta_1(m, n)$.

$$\text{Also} \qquad \int_0^1 f(x)\,dx = \frac{1}{\beta(m,n)} \int_0^1 x^{m-1}(1-x)^{n-1}\,dx$$

$$= \frac{\beta(m,n)}{\beta(m,n)} = 1, \text{ from (ii)}$$

m.g.f. with respect to the origin for this distribution cannot be expressed in a simple form.

Mean and Variance of this distribution :

$$\mu_1' = \int_0^1 x f(x)\,dx = \int_0^1 x \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}\,dx, \text{ from (i)}$$

$$= \frac{1}{\beta(m,n)} \int_0^1 x^m (1-x)^{m-1}\,dx$$

$$= \frac{1}{\beta(m,n)} \int_0^1 x^{(m+1)-1}(1-x)^{n-1}\,dx.$$

$$= \frac{1}{\beta(m,n)} \beta(m+1, n), \text{ from (ii)}$$

$$= \frac{\Gamma(m+n)}{\Gamma m \, \Gamma n} \cdot \frac{\Gamma(m+1)\,\Gamma n}{\Gamma(m+1+n)} = \frac{\Gamma(m+n)\, m\, \Gamma m}{\Gamma m\,(m+n)\,\Gamma(m+n)}$$

$\therefore \; \Gamma n = (n-1)\,\Gamma(n-1)$

or $\qquad \text{mean} = \mu_1' = \dfrac{m}{m+n}$ ...(iii)

$$\mu_2' = \int_0^1 x^2 f(x)\,dx = \frac{1}{\beta(m,n)} \int_0^1 x^2 . x^{m-1}(1-x)^{n-1}\,dx, \text{ from (i)}$$

$$= \frac{1}{\beta(m,n)} \int_0^1 x^{m+1}(1-x)^{n-1}\,dx$$

$$= \frac{1}{\beta(m,n)} \int_0^1 x^{(m+2)-1}(1-x)^{n-1}\,dx$$

$$= \frac{1}{\beta(m,n)} . \beta(m+2, n), \text{ from (ii)}$$

$$= \frac{\Gamma(m+n)}{\Gamma m \, \Gamma n} \cdot \frac{\Gamma(m+2)\,\Gamma n}{\Gamma(m+2+n)}, \text{ from (ii)}$$

$$= \frac{\Gamma(m+n)}{\Gamma m} \cdot \frac{(m+1)\, m\, \Gamma(m)}{(m+n+1)(m+n)\,\Gamma(m+n)},$$

$\because \; \Gamma n = (n-1)\,\Gamma(n-1)$

$$\text{or} \quad \mu_2' = \frac{m(m+1)}{(m+n)(m+n+1)}. \qquad \text{...(iv)}$$

$$\therefore \text{ Variance} = \mu_2 = \mu_2' - \mu_1'^2$$

$$= \frac{m(m+1)}{(m+n)(m+n+1)} - \left(\frac{m}{m+n}\right)^2, \text{ from (iii) and (iv)}$$

$$= \frac{m[(m+1)(m+n) - m(m+n+1)]}{(m+n)^2(m+n+1)}$$

$$= \frac{m[m(m+n) + (m+n) - m(m+n) - m]}{(m+n)^2(m+n+1)}$$

$$\frac{mn}{[(m+n)^2(m+n+1)]} \qquad \text{...(v)}$$

$$\text{And standard deviation } \sigma = \sqrt{\mu_2} = \frac{1}{(m+n)}\sqrt{\left[\frac{mn}{m+n+1}\right]} \qquad \text{...(vi)}$$

rth moment about origin of this distribution :

$$\mu_r' = \int_0^1 x^r f(x)\,dx = \frac{1}{\beta(m,n)}\int_0^1 x^r x^{m-1}(1-x)^{n-1}\,dx$$

$$= \frac{1}{\beta(m,n)}\int_0^1 x^{r+m-1}(1-x)^{n-1}\,dx$$

$$= \frac{\beta(r+m,n)}{\beta(m,n)}, \text{ from (ii)}$$

$$= \frac{\Gamma(r+m)\,\Gamma n}{\Gamma(r+m+n)} \cdot \frac{\Gamma(m+n)}{\Gamma m\,\Gamma n}, \text{ from (ii)}$$

$$= \frac{(r+m-1)!\,(m+n-1)!}{(r+m+n-1)!\,(m-1)!}, \quad \because \Gamma n = (n-1)!$$

☛ Q. 6. What do you understand by Normal Distribution ? Discuss normal distribution as a limiting form of Binomial Distribution.

Ans. The normal distribution was first discovered in 1733 by English mathematician De-Moivre, who obtained this continuous distribution as a limiting case of the binomial distribution and applied it to problems arising in the game of chance.

Throughout the eighteenth and nineteeth centuries, various efforts were made to establish the normal model as the underlying law ruling all continuous random variables. Thus, the name "*normal*". These efforts, however, failed because of false premises. The normal model has, nevertheless, become the most important probability model in satistical analysis.

Definition. *A random variable X is said to have a normal distribution with parameters μ (called "mean") and σ^2 (called "variance") if its density function is given by the probability law :*

$$f(x\,;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2\right]$$

or
$$f(x\,;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

$$-\infty < x < \infty,\ -\infty < \mu < \infty,\ \sigma > 0 \qquad \text{...(1)}$$

Notes : 1. A random variable X with mean μ and variance σ^2 and following the normal law (1) is expressed by $X \sim N(\mu, \sigma^2)$

2. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a standard normal variable with

$$E(Z) = 0 \text{ and } \mathrm{Var}(Z) = 1$$

and we write $Z \sim N(0, 1)$.

3. The p.d.f. of standard normal variate Z is given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z < \infty$$

and the corresponding distribution function, denoted by $\Phi(z)$ is given by

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{\infty} \varphi(u)\,du$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2}\,du$$

4. The graph of $f(x)$ is a famous '*bell-shaped*' curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

Normal Distribution as a Limiting form of Binomial Distribution.

Normal distribution is another limiting form of the binomial distribution under the following conditions :

(i) n, the number of trials is indefinitely large, *i.e.*, $n \to \infty$ and

(ii) neither p nor q is very small.

The probability function of the binomial distribution with parameters n and p is given by

$$p(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}\,; x = 0, 1, 2, \ldots, n \qquad \ldots(2)$$

Let us now consider the standard binomial variate :

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{npq}}\,; X = 0, 1, 2, \ldots, n \qquad \ldots(3)$$

When $\quad X = 0, Z = \dfrac{-np}{\sqrt{npq}} = -\sqrt{np/q}$

and when $\quad X = n, Z = \dfrac{n = np}{\sqrt{npq}} = \sqrt{nq/p}$

Thus in the limit as $n \to \infty$, Z takes the values from $-\infty$ to ∞. Hence the distribution of X will be a continuous distribution over the range $-\infty$ to ∞.

We want the limiting form of (2) under the above two conditions. Using Stirling's approximation to $r!$ for large r, *viz.*,

$$\lim_{r \to \infty} r! = \sqrt{2\pi}\, e^{-r} r^{r + (1/2)}.$$

we have in the limit as $n \to \infty$ and consequently $x \to \infty$.

$$\lim p(x) = \lim \left[\frac{\sqrt{2\pi}\, e^{-n} n^{n + \frac{1}{2}} p^x q^{n-x}}{\sqrt{2\pi}\, e^{-x} x^{x + \frac{1}{2}} \sqrt{2\pi}\, e^{-(n-x)} (n-x)^{n - x + \frac{1}{2}}} \right]$$

$$= \lim \left[\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} \cdot \frac{(np)^{x + \frac{1}{2}} (nq)^{n - x + \frac{1}{2}}}{x^{x + \frac{1}{2}} (n-x)^{n - x + \frac{1}{2}}} \right]$$

$$= \lim \left[\frac{1}{\sqrt{2\pi}\,\sqrt{npq}} \left(\frac{np}{x}\right)^{x + \frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n - x + \frac{1}{2}} \right] \qquad \ldots(4)$$

From (3), we have

$$X = np + Z\sqrt{npq}$$

$$\Rightarrow \qquad \frac{X}{np} = 1 + Z\sqrt{q/(np)}$$

Also

$$n - X = n - np - Z\sqrt{npq} = nq - Z\sqrt{npq}$$

$$\therefore \qquad \frac{n - X}{nq} = 1 - Z\sqrt{p/(nq)}$$

Also $\quad dz = \dfrac{1}{\sqrt{npq}}\, dx$

Hence the probability differential of the distribution of Z, in the limit is given from (4) by

$$d\,G(z) = g(z)\,dz = \lim_{n \to \infty} \left[\frac{1}{\sqrt{2\pi}} \times \frac{1}{N} \right] dz \qquad \ldots(5)$$

where $N = \left[\dfrac{x}{np}\right]^{x+\frac{1}{2}} \left[\dfrac{n-x}{nq}\right]^{n-x+\frac{1}{2}}$

$$\log N = (x+\tfrac{1}{2}) \log (x/np) + (n-x+\tfrac{1}{2}) \log \{(n-x)/nq\}$$

$$= (np + z\sqrt{npq} + \tfrac{1}{2}) \log [1 + z\sqrt{(q/np)}]$$

$$+ (nq - z\sqrt{npq} + \tfrac{1}{2}) \log [1 - z\sqrt{(p/nq)}]$$

$$= (np + z\sqrt{npq} + \tfrac{1}{2}) \left[z \cdot \sqrt{(q/np)} - \frac{1}{2} z^2 (q/np) + \frac{1}{3} z^3 (q/np)^{3/2} - \ldots\right]$$

$$+ (nq - z\sqrt{npq} + \tfrac{1}{2}) \left[- z\sqrt{(p/nq)} - \frac{1}{2} z^2 (p/nq) - \frac{1}{3} z^3 (p/nq)^{3/2} - \ldots\right]$$

$$= \left[\left\{z\sqrt{npq} - \frac{1}{2} qz^2 + \frac{1}{3} z^3 \frac{q^{3/2}}{\sqrt{np}} + z^2 q - \frac{1}{2} z^3 \frac{q^{3/2}}{\sqrt{np}} \right.\right.$$

$$\left. + \frac{1}{2} z \sqrt{q/np} - \frac{1}{4} z^2 \frac{q}{np} + \ldots \right)$$

$$+ \left(- z\sqrt{npq} - \frac{1}{2} z^2 p - \frac{1}{3} z^3 \frac{p^{3/2}}{\sqrt{nq}} + z^2 p \right.$$

$$\left.\left. + \frac{1}{2} z^3 \frac{p^{3/2}}{\sqrt{np}} - \frac{1}{2} z \sqrt{p/nq} - \frac{1}{4} z^2 \frac{p}{np} + \ldots \right\}\right]$$

i.e.,

$$\log N = \left[- \frac{1}{2} z^2 (p+q) + z^2 (p+q) + \frac{z}{2\sqrt{n}} \{\sqrt{q/p} + \sqrt{p/q}\} + 0\{n^{-1/2}\}\right]$$

$$= \frac{z^2}{2} + 0(n^{-1/2}) \rightarrow \frac{z^2}{2} \text{ as } n \rightarrow \infty$$

$$\therefore \lim_{n \to \infty} \log N = \frac{z^2}{2} \Rightarrow \lim_{n - \infty} N = e^{z^2/2}$$

Substituting in (5), we get

$$d\,G(z) = g(z)\,dz = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz, - \infty < z < \infty \qquad \ldots(6)$$

Hence the probability function of Z is

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, - \infty < z < \infty \qquad \ldots(7)$$

This is the probability density function of the ***normal distribution*** with mean 0 and unit variance.

If X is normal variate with mean μ and s.d. σ then $Z = (X - \mu)/\sigma$ is standard normal variate. Jacobian of transformation is

$1/\sigma$. Hence substituting in (7) the p.d.f. of a normal variate X with $E(X) = \mu$, Var $(X) = \sigma^2$ is given by

$$f_X(x) = \left\{\begin{array}{l} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty \\ 0, \textit{otherwise} \end{array}\right\}$$

Q.7. If x is normally distributed with zero mean and unit variance, find the expectation and variance of σ^2

Ans. The equation of the normal curve is

$$y = \frac{1}{\sigma\sqrt{(2\pi)}} e^{-(x-m)^2/2\sigma^2}$$

If mean is zero and variance is unit, then putting $m = 0$ and $\sigma = 1$, the above equation reduces to $y = \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}$

$\therefore$ Expectation (or mean) of x^2 *i.e.*, μ'_1

$$= \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}\, dx \qquad \textbf{(Note)}$$

$$= \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx \qquad ..(\text{i})$$

$$= -\frac{1}{\sqrt{(2\pi)}} \int x(-xe^{-x^2/2})\, dx \qquad \textbf{(Note)}$$

$$= -\frac{1}{\sqrt{(2\pi)}} \left[\left\{xe^{-x^2/2}\right\}_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 1.e^{-x^2/2}\, dx\right],$$

integrating by parts taking x as first function and remembering that $\frac{d}{dx}(e^{-x^2/2}) = -xe^{-x^2/2}$

$$= -\frac{1}{\sqrt{(2\pi)}} \left[0 - 2\int_0^{\infty} e^{-x^2/2}\, dx\right]$$

$$= \frac{2\sqrt{2}}{\sqrt{(2\pi)}} \int_0^{\infty} e^{-z^2}\, dz,$$

putting $\frac{x}{\sqrt{2}} = z$

$\therefore \quad dx = \sqrt{2}\, dz$

$$= \frac{2}{\sqrt{\pi}} \left(\frac{\sqrt{\pi}}{2}\right) = 1.$$

Also $$\mu_2' = \int_{-\infty}^{\infty} (x^2)^2 \cdot \frac{1}{\sqrt{(2\pi)}} e^{-x^2/2}\, dx \qquad \textbf{(Note)}$$

$$= -\frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} x^3 (-xe^{-x^2/2})\, dx \qquad \textbf{(Note)}$$

$$= -\frac{1}{\sqrt{(2\pi)}}\left[\left(x^3 e^{-x^2/2}\right)_{-\infty}^{\infty} - \int_{-\infty}^{\infty} 3x^2 . e^{-x^2/2}\, dx\right],$$

integrating by parts taking x^2 as first function.

$$= -\frac{1}{\sqrt{(2\pi)}}\left[0 - 3\int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx\right]$$

$$= 3\frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2}\, dx$$

$= 3\,(1)$, with the help of (i) above,

$\therefore$ Variance of $x^2 = \mu_2' = \mu_2{}^1 \mu_1'^2$

$= 3 - (1)^2 = 2$

☛ **Q. 8. In a normal distribution whose mean is 2 and S.D. 3, find a value of the variate such that the probability of the interval from the mean to the value is 0.4115. Find another value such that the probability for the interval from x = 3·5 to the value is 0.2307.**

Ans. Here mean $m' = 1$ and S.D. $= \sigma = 3$

$$\therefore \quad `t' = \frac{x - m}{\sigma} = \frac{x - 2}{3} \qquad ..(i)$$

Also from tables of the areas under the standard normal curve the value of t for which $\int_0^t \phi(t)\, dt = 0.4115$ is found to be 1.35.

$\therefore$ From (i), we get $\frac{1}{3}(x - 2) = 1.35$

or $\qquad x - 2 = 4.05$

or $\qquad x = 6.05$

Again for $x = 3.5$ from (i), we have $t = \frac{3.5 - 2}{3} = 0.5.$

Now it is given that the area under the standard normal curve between $t = 0.5$ to $t = t_1$ (say) is 0.2307. Also from the tables of the areas under the standard normal curves we find that the area between $t = 0$ to $t = 0.5$ is 0.1915.

$\therefore$ Area between $t = 0$, to $t = t_1$ is $0.1915 + 0.2307 = 0.4222.$

$\therefore$ From the tables of areas under the standard normal curve the value of $t_1 = 1.42.$

Hence from (i), we have $\frac{x - 2}{3} = 1.42$

or $\qquad x - 2 = 4.26$

or $\qquad x = 6.26$

Q. 9. In a normal distribution 31% of items are under 45 and 8% are over 64. Find the mean and standard deviation of the distribution.

Ans. Let m and σ be the mean and standard deviation respectively of the distribution.

Given that 31% of items are under 45, so the percentage of items lying between 45 and mean m is 50–31 *i.e.*, 19.

∴ Using usual notations

We have $$\int_{(45-m)/\sigma}^{0} \phi(t)\,dt = \frac{19}{100} = 0.19,$$

where $t = \dfrac{x-m}{\sigma}$

or $$\int_{0}^{(m-15)/\sigma} \phi(t)\,dt = 0.19 \qquad \textbf{(Note)}$$

This gives, from the table of areas under the normal curve,

$$\frac{m-45}{\sigma} = 0.5 \text{ nearly (from the table)}$$

i.e., $$m - 45 = (0.5)\,\sigma \qquad ..(i)$$

Similarly it is given that 8% of items are over 64, so the percentage of items lying between mean m and 64 is 50–8 *i.e.*, 42.

∴ Using above notations, we have

$$\int_{0}^{(66-m)/\sigma} \phi(t)\,dt = \frac{42}{100} = 0.42$$

which give

$\dfrac{64-m}{\sigma} = 1.4$, from the table of areas under the normal curve.

i.e., $$64 - m = (1.4)\,\sigma \qquad ..(ii)$$

Adding (i) and (ii) we get $19 = 1.9\,\sigma$ or $\sigma = 10$

∴ From (i), we get $m = 45 + (0.5) \times 10 = 45 + 5 = 50.$ ●

Q.10. For the Beta variate of the first kind with parameters *m* and *n*, evaluate Harmonic Mean.

Ans. We know that the probability density function of the Beta variate of the first kind is given by

$$f(x) = \frac{1}{\beta(m,n)\,x^{n-1}\,(1-x)^{m-1}},$$
$$0 \le x \le 1;\ m, n > 0 \qquad ..(i)$$

If H be the harmonic mean of this distribution, then

$$\frac{1}{H} = E\left(\frac{1}{X}\right) \int_0^1 \frac{1}{x}(x)\,dx, \text{ by definition}$$

$$= \int_0^1 \frac{1}{x} \cdot \frac{1x^{m-1}}{\beta(m,n)} (1-x)^{n-1}\,dx$$

$$= \frac{1}{\beta(m,n)} \int_0^x (m-1)^{-1} (1-x)^{n-1}\, dx$$

$$= \frac{1}{\beta(m,n)} \beta(m-1, n)$$

$$= \frac{\Gamma(m+n)}{\Gamma m \,\Gamma n} \cdot \frac{\Gamma(m-1)\,\Gamma n}{\Gamma(m-1+n)}$$

$$= \frac{(m+n-1)\,\Gamma(m+n-1)}{(m-1)\,\Gamma(m-1)} \cdot \frac{\Gamma(m-1)}{\Gamma(m-1+n)}$$

$$= \frac{m+n-1}{m-1}$$

$$\therefore \quad H = \frac{m-1}{m+n-1}$$

☛ **Q.11. Write about Log-normal Distribution.**

Ans. Log-normal Distribution : The positive *r.v.* X is said to have a log normal distribution if $\log_e X$ is normally distributed.

Let $Y = \log_e X \sim N(\mu, \sigma^2)$. For $x > 0$,

$$F_X(x) = P(X \le x) = P(\log_e X \le \log_e x) = P(Y \le \log_e x)$$

(Since $\log X$ is monotonic increasing function.)

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log x} \exp\{-(y-\mu)^2/\sigma^2\}\, dy$$

[Since $Y \sim N(\mu, \sigma^2)$]

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_0^x \exp\{-(\log u - \mu)^2/2\sigma^2\} \frac{du}{u},$$

$(y = \log u)$

For $x \le 0$, $F_X(x) = P(X \le x) = 0$, because X is a positive *r.v.*

Let us define

$$f_X(u) = \begin{cases} \dfrac{1}{u\sigma\sqrt{2\pi}} \cdot \exp\{-(\log u - \mu)^2/2\sigma^2\}, & u > 0 \\ 0, & u \le 0 \end{cases} \quad ..(i)$$

Then $F_X(x) = \int_{-\infty}^{x} f_X(u)\, du$,

for every x and hence $f(x)$ defined in (1) is a *p.d.f.* of X.

Remark : If $X \sim N(\mu, \sigma^2)$, then $Y = e^x$, is called a log-normal random variable, since its logarithm $\log Y = X$, is a normal *r.v.*

Moments : The rth moment about origin is given by :

$$\mu_r' = E(X^r) = E(e^{rY}) \qquad [\because Y = \log X \Rightarrow X = e^Y]$$

$$= M_Y(r) \qquad \text{(m.g.f. of } Y, r \text{ being the parameter)}$$

$$= \exp\left(\mu r + \frac{1}{2} r^2 \sigma^2\right) \qquad [\because Y \sim N(\mu, \sigma^2)] \quad ..(2)$$

Remarks 1. In particular if we take $\mu = \log \alpha, \alpha > 0$, *i.e.*, $\log X \sim N(\log \alpha, \sigma^2)$, then

$$\mu_r' = E(X^r) = \exp\{r . \log \alpha + \frac{1}{2} r^2 \sigma^2\} = \alpha^r . \exp\{r^2 \sigma^2 / 2\} \qquad ..(3)$$

$$\therefore \quad \text{Mean} = \mu_1' = \alpha e^{\sigma^2/2}$$

and $$\mu_2 = \mu_2' - \mu_1'^2 = \alpha^2 e^{\sigma^2} (e^{\sigma^2} - 1)$$

2. Log normal distribution arises in problems of economics, biology, geology and reliability theory. In particular, it arises in the study of dimensions of particles under pulverisation.

3. If $X_1, X_2, \ldots\ldots\ldots X_n$ is a set of independently identically distributed random variables such that mean of each $\log X_i$ is μ and its variance is σ^2, then the product $X_1 . X_2 \ldots\ldots X_n$ is asymptotically distributed according to logarithmic normal distribution and with mean μ and variance $n\sigma^2$.

Q.12. Write about Standard Laplace (Double Exponential) Distribution.

Ans. **Standard Laplace (Double Exponential) Distribution**

Definition : A continuous *r.v.* X is said to have standard Laplace (double exponential distribution if its *p.d.f.* is given by :

$$f(x) = \frac{1}{2} \exp. (- |x|);$$
$$- \infty < x < \infty. \qquad ..(1)$$

Characteristic Function of Standard Laplace Distribution :

$$\varphi_x(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx = \frac{1}{2} \int_{-\infty}^{\infty} e^{itx} e^{-|x|}\, dx$$

$$= \frac{1}{2} \left[\int_{-\infty}^{\infty} \cos tx . e^{-|x|}\, dx + i \int_{-\infty}^{\infty} \sin tx\, e^{-|x|}\, dx \right]$$

$$= \frac{1}{2} . 2 \int_0^{\infty} \cos tx . e^{-|x|}\, dx$$

since the integrands in the first and second integrals are even and odd function of x respectively.

$$\therefore \quad \varphi_x(t) = \int_0^{\infty} e^{-x} \cos tx\, dx = 1 - t^2 \int_0^{\infty} e^{-x} \cos tx\, dx$$

(On integration by parts)

$$= 1 - t^2 \varphi_x(t)$$

$$\Rightarrow \quad \varphi_x(t) = \frac{1}{1 + t^2} \qquad ..(2)$$

$$\therefore \quad \phi_x(t) = 1 - t^2 + t^4 - t^6 + \ldots\ldots\ldots\ldots$$

$$= 1 + 2 . \frac{(it)^2}{2!} + 4! \frac{(it)^4}{4!} - .....$$

Hence $\kappa_1 = \kappa_3 = 0;$

$\kappa_2 = 2; \kappa_4 = 4! = 24$

$\therefore$ Mean $= \kappa_1 = 0;$

Variance $= \mu_2 = \kappa_2 = 2$

$\mu_3 = \kappa_3 = 0;$

$\mu_4 = \kappa_4 + 3k_2^2 = 24 + 12 = 36$

$$\Rightarrow \qquad \beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0;$$

$$\beta = \frac{\mu_4}{\mu_2^2} = 9$$

Remark : Men deviation about mean for standard Laplace distribution is 1. (Try it).

Two Parameter Laplace Distribution

Definition : A continuous *r.v.* X is said to have a double exponential (Laplace) distribution with two parameters λ and μ if its p.d.f. is given by :

$$f(x); \mu, \sigma = \frac{1}{2\lambda} \exp(- | x - \mu |/\lambda);$$

$$- \infty < x < \infty, \lambda > 0$$

We write $X \sim$ Lap (λ, μ)

Let $Y = \frac{x - \mu}{\lambda} \Rightarrow X = \mu + \lambda Y$

The *p.d.f.g* (.) of Y is given by :

$$g(y) = f(x) \left| \frac{dx}{dy} \right| = \frac{1}{2\lambda} \exp. (- | y |) . \lambda \qquad \text{[From (e)]}$$

$$= \frac{1}{2} \exp(- | y |);$$

$$- \infty < y < \infty,$$

which is the *p.d.f.* of standard Laplace distribution

Hence, if $X \sim$ Lap (λ, μ),

Then $$Y = \frac{X - \mu}{\lambda} \sim \text{Lap}(1, 0)$$

i.e., Y has standard Laplace distribution.

Characteristic Function of Two Parameter Laplace Distribution : If $X \sim$ Lap (λ, μ), then

$$\phi_X(t) = E(e^{itX}) = E[e^{it(\mu + \lambda Y)}],$$

where $$Y = \frac{X - \mu}{\lambda} \sim \text{Lap}(1, 0)$$

$$= e^{it\mu} \cdot E\left(e^{i\lambda t . Y}\right) = e^{it\mu} \cdot \phi_Y(\lambda t)$$

$$= \frac{e^{it\mu}}{1 + \lambda^2 t^2}$$

$$\left[\because \text{Y is standard Laplace variate with } \phi(t) = \frac{1}{1+t^2}\right]$$

Moments of Two Parameter Laplace Distribution : If $X \sim$ Lap (λ, μ), the rth moment about origin is given by :

$$\mu_r' = E(X^r) = \frac{1}{2\lambda}\int_{-\infty}^{\infty} x^r \exp\left(\frac{-|x-\mu|}{\lambda}\right) dx$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} (z\lambda + \mu)^r \exp(-|z|)\, dz, \quad \left(z = \frac{x-\mu}{\lambda}\right)$$

$$= \frac{1}{2}\int_{-\infty}^{\infty} \left[\sum_{k=0}^{r} \binom{r}{k} (z\lambda)^k \mu^{r-k}\right] \exp(-|z|)\, dz$$

$$= \frac{1}{2}\sum_{k=0}^{r} \left[\binom{r}{k} \lambda^k \mu^{r-k} \int_{-\infty}^{\infty} z^k \exp(-|z|)\, dz\right]$$

$$\frac{1}{2}\sum_{k=0}^{r} \left[\binom{r}{k} \lambda^k \mu^{r-k} \left\{\int_{-\infty}^{0} z^k e^{-|z|}\, dz + \int_{0}^{\infty} z^k e^{-|z|}\, dz\right\}\right]$$

$$= \frac{1}{2}\sum_{k=0}^{r} \left[\binom{r}{k} \lambda^k \mu^{r-k} \left\{(-1)^k \int_{0}^{\infty} e^{-z} z^k\, dz + \int_{0}^{\infty} e^{-z} z^k\, dz\right\}\right]$$

$$= \frac{1}{2}\sum_{k=0}^{r} \left[\binom{r}{k} \lambda^k \mu^{r-k} \Gamma(k+1)\{(-1)^k + 1\}\right]$$

$$\Rightarrow \mu_r' = \frac{1}{2}\sum_{k=0}^{r} \left[\binom{r}{k} \lambda^k \mu^{r-k} k!\{1 + (-1)^k\}\right]$$

$\therefore$ Mean $= \mu_1' = \mu, \mu_2' = \mu^2 + 2\lambda^2$

and Variance $= \mu_2 = \mu_2' - \mu_1'^2 = 2\lambda^2$.

Similarly we can obtain higher order moments from (iii) and hence the values of β_1 and β_2 can be obtained.

Q.13. Write about the Weibul Distribution.

Ans. A random variable X has a Weibul distribution with three parameters $c\ (> 0)\ \alpha\ (> 0)$ and μ if the $r.v.$ $Y = \left(\frac{X-\mu}{\sigma}\right)^c$ (i) has the exponential distribution with $p.d.f.$

$$p_Y(y) = e^{-y}, y > 0 \qquad \text{(ia)}$$

Definition : A continuous $r.v.$ X has a Weibul distribution with parameters $c\ (> 0)$, $a\ (> 0)$ and μ if its p.d.f. is :

$$f(x; c, \alpha, \mu) = \frac{c}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{c-1}$$

$$\exp\left\{-\left(\frac{x-\mu}{\alpha}\right)^{c}\right\}; x > ,i, c > 0. \qquad ..(1b)$$

The standard Weibul distribution is obtained on taking $\alpha = 1$ and $\mu = 0$, so that the *p.d.f.* of standard Weibul distribution which depends only on a single parameter c is given by :

$$px(x) = c\,x^{c-1} . \exp(-x^{c}); \qquad ...(1c)$$

$$x > 0, c > 0$$

Moments of Standard Weibul Distribution : For standard Weibul distribution , $(\alpha = 1, \mu = 0)$, from (i), we get $Y = X^c$ which has the exponential distribution (ia). We have

$$\mu_r' = E(X^r) = E(Y^{1/c})^r = E(Y^{r/c}) = \int_0^\infty e^{-y} y^{r/c}\,dy$$

$$\Rightarrow \mu_r' = \Gamma\left(\frac{r}{c} + 1\right)$$

$$\therefore \text{ Mean} = E(X) = \Gamma\left(\frac{1}{c} + 1\right)$$

and

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= \Gamma\left(\frac{2}{c} + 1\right) - \left[\Gamma\left(\frac{1}{c} + 1\right)\right]^2$$

Similarly, we can obtain expressions for higher order moments and hence for β_1 and β_2. For large c, the mean is approximated by :

$$E(X) - 1 - \frac{\gamma}{c} + \frac{1}{2c^2}\left(\frac{\pi^2}{6} + \gamma^2\right)$$

$$= 1 - 0.57722\,c^{-1} + 0.98905\,c^{-2},$$

where $\gamma = 0.57722$ is Euler's constant.

The distribution is named after Waloddi Weibul, a Swedish physicist, who used it in 1939 to represent the distribution of the breaking strength of materials. Kao, J.H.K. (1958-59) advocated the use of this distribution in reliability studies and quality control work. It is also used as a tolerance distribution in the analysis of quantum response data.

Characterisation of Weibul Distribution : Dubey, S.D. (1968) has obtained the following result :

"Let X_i $(i = 1, 2, ..., n)$ be *i.i.d.* random variables. Then min $(X_1, X_2, ..., X_n)$ has a Weibul distribution if and only if the commondistribution of $X_i's$ is a Weibul distribuiton.

Proof . Let X_i, $(i = 1, 2, ..., n)$ be *i.i.d. r.v's.* each with Weibul distribution (9.28b) and let $Y = \min(X_1, X_2, ... X_n)$. Then

$$P(Y > y) = P\{\min(x_1, X_2, ..., X_n) > y\}$$

$$= P\bigcap_{i=1}^{n} X_i > y)$$

$$= \prod_{i=1}^{n} P(X_1 > y) = [P(X_i > y)]^n,$$

(since X_i' s are *i.i.d.r.v.'s*)

Now $P(X_i > y) = \int_y^\infty c\,\alpha^{-1}\left(\frac{x-\mu}{\alpha}\right)^{c-1}\exp\left\{-\left(\frac{x-\mu}{\alpha}\right)^c\right\}dx.$

$$= \int_{[(y-\mu/\alpha)]^c}^\infty e^{-t}\,dt,\ t = \left(\frac{x-\mu}{\alpha}\right)^c$$

$$= \exp\left\{-\left(\frac{y-\mu}{\alpha}\right)^c\right\}^c$$

Substituting in (*), we get

$$P(Y > y) = \left[\exp\left\{-\left(\frac{y-\mu}{\alpha}\right)^c\right\}\right]^n$$

$$= \exp\left\{-n\left(\frac{y-\mu}{\alpha}\right)^c\right\} = \exp\left[-\left\{\frac{n^{1/c}(y-\mu)}{\alpha}\right\}^c\right]$$

This implies that Y has the same Weibul distribution as X_i 's with the difference that the parameter α is replaced by $\alpha n^{-1/c}$. ●

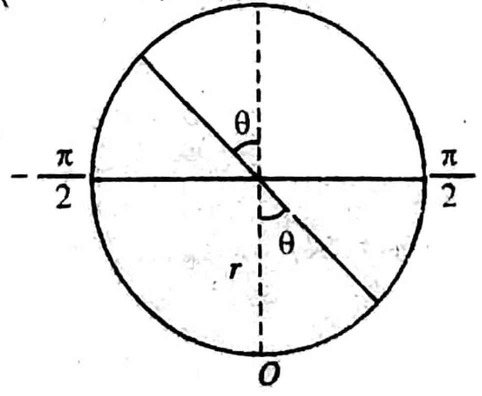
Q.14. What is Cauchy Distribution and Write Characteristics of it.

Ans.

Cauchy Distribution

Let us consider a roulette wheel in which the probability of the pointer stopping at any part of the circumference is constant. In other words, the probability that any value of θ lies in the interval $[-\pi/2, \pi/2]$ is constant and consequently θ is a rectangular variate in the range $[-\pi/2, \pi/2]$ with probability differential given by :

$$dP(\theta) = \begin{cases} (1/\pi)\,d\theta, & -\pi/2 \le \theta \le \pi/2 \\ 0, \text{otherwise} \end{cases} \quad \text{...(i)}$$



Let us now transform to variable X by the substitution :

$x = r\tan\theta \Rightarrow dx = sec^2\theta\, d\theta$

Since, $-\pi/2 \le \theta \le \pi/2$, the range for X is from $-\infty$ to ∞. Thus the probability differential of X becomes :

$$dF(x) = \frac{1}{\pi}\cdot\frac{dx}{r\sec^2\theta} = \frac{1}{\pi}\cdot\frac{dx}{[r\{1+(x^2/r^2)\}]}$$

$$= \frac{r}{\pi}\cdot\frac{dx}{r^2+x^2};\ -\infty < x < \infty$$

In particular if we take $r = 1$, we get : $f(x)\,\frac{1}{\pi}\cdot\frac{1}{1+x^2}$, $-\infty \angle x < \infty$

This is the p.d.f. of a standard Cauchy variate and we write $X \sim C(1, 0)$.

Definition : A random variable X is said to have a standard cauchy distribution if its p.d.f. is given by :

$$f_X(x) = \frac{1}{\pi(1+x^2)},\ -\infty < x < \infty \qquad ...(ii)$$

and X is termed as standard Cauchy variate.

More generally, Cauchy distribution with parameters λ and μ has the p.d.f.,

$$g_y(y) = \frac{\lambda}{\pi[\lambda^2+(y-\mu)^2]},\ -\infty < y < \infty; \lambda > 0 \qquad ...(iia)$$

and we write $X \sim C(\lambda, \mu)$

But putting $X = (Y-\mu)/\lambda$ in (iia), we get (ii).

Hence if $Y \sim C(\lambda, \mu)$,

then $\qquad X = (Y-\mu)/\lambda \sim C(1, 0) \qquad ...(iib)$

Characteristic Function of (Standard) Cauchy Distribution : If X is a standard Cauchy variate than

$$\varphi_x(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{itx}}{1+x^2}\,dx \qquad ..(*)$$

To evaluate (*) consider standard Laplace distribution $f_1(z) = \frac{1}{2}e^{-|z|}, -\infty < z < \infty$.

Then $\qquad \varphi_i(t) = \phi_Z(t) = E(e^{itZ}) = \frac{1}{1+t^2}$

Since $\varphi_1(t)$ is absolutely integrable in $(-\infty, \infty)$, we have by Inversion theorem

$$\frac{1}{2} e^{-|z|} = f_1(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itz} \varphi_1(t)\, dt$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itz}}{1+t^2}\, dt$$

$$\Rightarrow \quad e^{-|z|} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itz}}{1+t^2}\, dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itz}}{1+t^2}\, dt$$

[Changing t to $-t$]

On interchanging t and z, we have $e^{-|t|} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{itz}}{1+z^2}\, dz$ (**)

From (*) and (**), we get

$$\varphi_X(t) = e^{-|t|}$$

Remarks 1. If Y is a Cauchy variate with parameters λ and μ, then,

$$X = \frac{Y-\mu}{\lambda} \sim C(1, 0)$$

$$\Rightarrow \quad Y = \mu + \lambda X$$

$$\therefore \quad \phi_Y(t) = E(e^{itY}) = e^{i\mu t} E(e^{it\lambda}) = e^{i\mu t} \phi_X(t\lambda)$$

$$= e^{i\mu t - \lambda|t|}, \lambda > 0$$

2. Additive Property of Cauchy Distribution : If X_1 and X_2 are independent Cauchy variates with parameters. (λ_1, μ_1) and (λ_1, μ_2) respectively, then $X_1 + X_2$ is a Cauchy variate with parameters $(\lambda_1 + \lambda_2, \mu_1 + \mu_2)$.

Proof : $\phi_{xj} = \exp\ \{i\mu_j t - \lambda_j\ |t|\ \}, (j = 1, 2)$ [From (a)]

$$\therefore \quad \phi_{x_i} + x_2(t) = \phi_{x_i}(t)\, \phi_{x_2}(t)$$

(Since X_1, X_2 are independently)

$$= \exp\ [it\ (\mu_1 + \mu_2) - (\lambda_1 + \lambda_2)\ |t|\ ]$$

and the result follows by uniqueness theorem of characteristic functions.

3. Since $\phi' x(t)$ in (1) [where (') denotes differentiation w.r.t to t] does not exist at $t = 0$, the mean of the Cauchy distribution does not exist.

4. Let $X_1, X_2, \ldots, X_n$ be a sample of n independent observations from a standard Cauchy distribution and define $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. Then

$$\phi_{\bar{X}}(t) = \phi_{\Sigma x_j}(t/n) = \prod_{j=1}^{n}[\phi_{x_j}(t/n)] = [\phi_{x_j}(t/n)]^n$$

(since X_j' s are *i.i.d.*)

$$= [e^{-1|t/n|}]^n = e^{-|t|} = \phi_x(t)$$

Hence by uniqueness theorem of characteristic functions, we have :

"The arithmetic mean $\bar{X}$ of sample $X_1, X_2, \ldots, X_n$ of independent observations from a standard Cauchy distribution is also a standard Cauchy variate.

In other words, the arithmetic mean of a random sample of any size yields exactly as much information as a single determination of X."

The implies that the sample mean $\bar{X}$„ of random sample of size n, as an estimate of population mean does not improve with increasing n, which contradicts in Weak Law of Large Numbers (WLLN).

Q. 15. Define exponential distribution and obtain its mean and variance.

Ans. For this distribution the probability density function is given by $f(x) = ce^{-x/\theta}, \theta > 0$ and the range of the exponential variate is $0 \le x \le \infty$

Since it is a probability distribution, so

$$1 = \int_0^\infty f(x)\,dx = \int_0^\infty ce^{-x/\theta}\,dx = \left[ce^{-x/\theta}(-\theta)\right]_0^\infty$$

or1 $= 0 - (-c\theta) = c\theta$, which gives $c = 1/\theta$.

∴ The probability density function of an exponential variate reduces to

$$f(x) = (1/\theta)\,e^{-x/\theta}, \theta > 0, 0 \le x \le \infty$$

m.g.f for Exponential Distribution

$$M_0(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x)\,dx = \int_0^\infty e^{tx}\left(\frac{1}{\theta}e^{-x/\theta}\right)dx$$

$$= \frac{1}{\theta}\int_0^\infty e^{-x[1/\theta)-t]}\,dx = \frac{1}{\theta}\left[\frac{e^{-x[1/\theta)-t]}}{-\{(1/\theta)-t\}}\right]_0^\infty$$

$$= \frac{1}{\theta}\left[0 + \frac{1}{(1/\theta)-t}\right] = \frac{1}{\theta} \times \frac{\theta}{1-\theta t} = (1-\theta t)^{-1}$$

Q. 16. Obtain recurrence relation for finding moments of Poisson distribution. What is the additive property of independent Poisson variates ? Prove it.

Ans. $\mu_1' = E(X) \sum_{x=0}^{\infty} x\,p(x,\lambda) = \sum_{x=0}^{\infty} x \,.\, \frac{e^{-\lambda}\lambda^x}{x!}$

$$= \lambda e^{-\lambda} \left\{ \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \right\}$$

$$= \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots \right) = \lambda e^{-\lambda} \,.\, e^{\lambda} = \lambda$$

Hence the mean of the Poisson distribution is λ.

$$\mu_2' = E(X^2) = \sum_{x=0}^{\infty} x^2 p(x,\lambda)$$

$$= \sum_{x=0}^{\infty} \{x(x-1) + x\} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda} \sum_{x=0}^{\infty} x(x-1) \frac{\lambda^x}{x!} + \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \lambda^2 e^{-\lambda} \left[\sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \right] + \lambda$$

$$= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda = \lambda^2 + \lambda$$

$$\mu_3' = E(X)^3 = \sum_{x=0}^{\infty} x^3 p(x,\lambda) \sum_{x=0}^{\infty}$$

$$\{x(x-1)(x-2) + 3x(x-1) + x\} \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= \sum_{x=0}^{\infty} x(x-1)(x-2) \frac{e^{-\lambda}\lambda^x}{x!} + 3 \sum_{x=0}^{\infty} x(x-1) \frac{e^{-\lambda}\lambda^x}{x!}$$

$$+ \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!}$$

$$= e^{-\lambda}\lambda^3 \left\{ \sum_{x=3}^{\infty} \frac{\lambda^{x-3}}{(x-3)!} \right\} + 3e^{-\lambda}\lambda^2$$

$$\left\{ \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} \right\} + \lambda$$

$$= e^{-\lambda}\lambda^3 e^{\lambda} + 3e^{-\lambda}\lambda^2 e^{\lambda} + \lambda = \lambda^3 + 3\lambda^2 + \lambda$$

Recurrence Relation for Moments of the Poisson Distribution :

By def.

$$\mu_r = E\{X - E(X)\}^r = \sum_{x=0}^{\infty} (x-\lambda)^r p(x, \lambda)$$

$$= \sum_{x=0}^{\infty} (x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!}$$

Differentiating *w.r.to* λ, we get

$$\frac{d\mu_r}{d\lambda} = \sum_{x=0}^{\infty} r(x-\lambda)^{r-1}(-1)\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{x!}\{x\lambda^{x-1}e^{-\lambda} - \lambda^x e^{-\lambda}\}$$

$$= -r\sum_{x=0}^{\infty} (x-\lambda^{r-1} . \frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty} \frac{(x-\lambda)^r}{x!}\{\lambda^{x-1}.e^{-\lambda}(x-\lambda)\}$$

$$= -r\sum_{x=0}^{\infty} (x-\lambda)^{r-1} . \frac{e^{-\lambda}\lambda^x}{x!} + \frac{1}{\lambda}\sum_{x=0}^{\infty} (x-\lambda)^{r+1} . \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\therefore \quad = -r\mu_{r-1} + \frac{1}{\lambda}\mu_{r+1}$$

$$\Rightarrow \quad \mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}$$

Putting $r = 1, 2$ and 3 successively, we get

$$\mu_2 = \lambda\mu_0 + \lambda\frac{d\mu_1}{d\lambda} = \lambda,$$

$$\mu_3 = 2\lambda\mu_1 + \lambda\frac{d\mu_2}{d\lambda} = \lambda,$$

$$\mu_4 = 34\lambda\mu_2 + \lambda\frac{d\mu_3}{d\lambda} = 3\lambda^2 + \lambda$$

Additive or Reproductive Property of Independent Poisson Variates : Sum of independent Poisson variates is also a Poisson variate. More elaborately, if X_i; ($i = 1, 2, \ldots\ldots, n$) are independent Poisson variates with parameters $\lambda_i, i = 1, 2, \ldots\ldots\ldots, n$ respectively, then $\sum_{i=1}^{n} X_i$ is also a Poisson variate with parameter $\sum_{i=1}^{n} \lambda_i$.

Proof : $M_{X_i}(t) = e^{\lambda_i(e^t - 1)}$; $i = 1, 2, \ldots\ldots, n$ $\quad [\because X_i \sim P(\lambda_i)]$

$$M_{X_1 + X_2 + \ldots\ldots + X_n}(t) = M_{X_1}(t)\, M_{X_2}(t) \ldots\ldots M_{X_n}(t),$$

[Since $X_i; i = 1, 2, \ldots\ldots, n$ are independent.]

$$= e^{\lambda_1 (e^t - 1)} e^{\lambda_2 (e^t - 1)} \ldots\ldots e^{\lambda_n (e^t - 1)} = e^{(\lambda_1 + \lambda_2 + \ldots\ldots + \lambda_n)(e^t - 1)}$$

which is the *m.g.f.* of a Poisson variate with parameter $\lambda_1 + \lambda_2 + \ldots\ldots + \lambda_n$. Hence, by uniqueness theorem of *m.g.f.'s*, $\sum_{x=1}^{n} X_i$ also a Poisson variate with parameter $\sum_{x=1}^{n} \lambda_i$.

●

Q. 17. Define negative binomial distribution Find its m.g.f. and hence find its mean and variance.

Ans.

Negative Binomial Distribution

The equality of the mean and variance is an important characteristic of the Poisson distribution, whereas for the binomial distribution the mean is always greater than the variance. Occasionally, However, observable phenomena give rise to empirical discrete distributions which show a variance larger than the mean. Some of the commonest examples of such behaviour are the frequency distributions of plant density obtained by quadrant sampling when the clustering of plants makes the simple Poisson model inapplicable. It has been shown by different investigators that in such cases the negative binomial distribution provides an excellent model because this distribution has a variance larger than the mean. Bacterial clustering (or contagion), *e.g.*, deaths of insects, number of insect bites leads to the negative binomial distribution and the distribution also arises in inverse sampling from a binomial population or as a weighted average of Poisson distribution. This important probability distribution is sometimes also referred to as the Pascal distribution after the French mathematician Blaise Pascal (1623-62), but there seems to be no historical justification. The negative binomial distribution can be derived from empirical considerations in many ways. Here we consider the Binomial probability situation with some modifications.

Suppose we have a succession of n Bernoulli trials. We assume that (*i*) the trials are independent, (ii) the probability of succes 'p' in a trial remains constant from trial to trial.

Let $f(x; r, p)$ denote the probability that there are x failures preceding the rth success in $x + r$ trials. Now, the last trial must be a success, whose probability is p. In the remaining $(x + r - 1)$ trials we must have $(r - 1)$ successes whose probability is given by binomial probability law by the expression : $\binom{x + r - 1}{r - 1} p^{r-1} q^x$

Therefore, by compound probability theorem, $f(x; r, p)$ is given by the product of these two probabilities, *i.e.*,

$$f(x, r, p) = \binom{x+r-1}{r-1} p^{r-1} q^{x} \cdot p = \binom{x+r-1}{r-1} p^{r} q^{x}$$

Definition : A random variable X is said to follow a negative binomial distribution with parameters r and p if its probability mass function is given by :

$$P(X = x) = p(x) = \begin{cases} \binom{x+r-1}{r-1} p^{r} q^{x}; x = 0, 1, 2, \ldots \\ 0, \text{ otherwise} \end{cases} \qquad \ldots(i)$$

Moment Generating Function of Negative Binomial Distribution :

$$M_X(t) = E(e^{tx}) = \sum_{x=0}^{\infty} e^{tx} p(x) = \sum_{x=0}^{\infty} \binom{-r}{x} Q^{-r} \left(-\frac{Pe^{t}}{Q}\right)^{x} \qquad \ldots(ii)$$

$$= (Q - Pe^{t})^{-r}$$

$$\mu_1' = \left[\frac{d}{dt} M(t)\right]_{t=0} = [-r(-Pe^{t})(Q - Pe^{t})^{-r-1}]_{t=0} = rP$$

$\therefore$ Mean of the negative binomial distribution is rP. ...(iia)

Q. 18. Discuss the main features of a normal distribution.

Ans. Chief Characteristics of the Normal Distribution and Normal Probability Curve : The normal probability curve with mean μ and standard deviation #is given by the equation :

$$f(x) \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-mu)^2/2\sigma^2}, -\infty < x < \infty$$

and has the follows properties :

(i) The curve is bell-shaped and symmetrical about the line $x = \mu$.

(ii) Mean, median and mode of the distribution coincide.

(iii) As x increases numerically, $f(x)$ decreases rapidly, the maximum probability occurring at the pont $x = \mu$, and is given by :

$$[p(x)]_{\max} = \frac{1}{\sigma\sqrt{2\pi}}$$

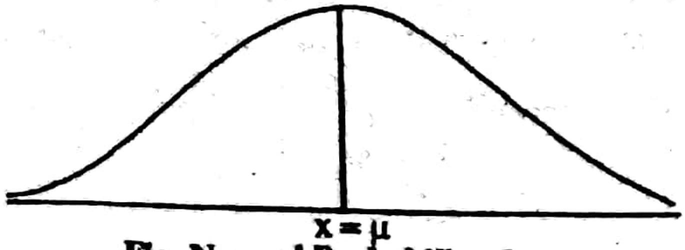


Fig. Normal Probability Curve

(iv) $\beta_1 = 0$ and $\beta_2 = 3$

(v) $\mu_{2r+1} = 0, (r = 0, 1, 2, \ldots)$, and

$$\mu_{2r} = 1.3.5 \ldots (2r-1)\,\sigma^{2r}, (r = 0, 1, 2, \ldots)$$

(vi) Since $f(x)$ being the probability, can never be negative, no portion of the curve lies below the x-axis.

(vii) Linear combination of independent normal variates is also normal variate.

(viii) x-axis is an asymptote to the curve.

(ix) The points of inflexion of the curve are :

$$x = \mu \pm \sigma, f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2}$$

(x) Mean deviation about mean $= \sqrt{\frac{2}{\pi}}\,\sigma \sim \frac{4}{5}\sigma$ (approx.)

(xi) Quartiles are given by :

$$Q_1 = \mu - 0{\cdot}6745\,\sigma; \quad Q_3 = \mu + 0{\cdot}6745\,\sigma$$

(xii) $Q.D. = \frac{Q_3 - Q_1}{2} \sim \frac{2}{3}\sigma$. We have (approximately)

$$Q:d:M.D.:SD. :: \frac{2}{34}\sigma : \frac{4}{5}\sigma : \sigma :: \frac{2}{3} : \frac{4}{5} : 1$$

$\Rightarrow$ $Q.D. : M.D. : s.D. :: 10 : 12 : 15$

(xiii) Area Propertry :

$P(\mu - \sigma < X < \mu + \sigma) = 0{\cdot}6826$,

$P(\mu - 2\sigma < X \mu + 2\sigma) = 0{\cdot}9544$

and $P(\mu - 3\sigma < X < \mu + 3\sigma) = 0{\cdot}9973$

Q. 19. For a rectangular distribution :

$$f(x) = \frac{1}{2a}, \; -a \le x \le a$$

Show that : $\mu_{2r} = \dfrac{a^{2r}}{2r+1}$

Ans. Here $M_0(t) = \int_{-a}^{a} e^{tx} \cdot \frac{1}{2a}\,dx$, by def. **(Note)**

$$= \frac{1}{2a}\left(\frac{e^{tx}}{t}\right)_{-a}^{a} = \frac{1}{2at}\left(e^{at} - e^{-at}\right)$$

$$= \frac{1}{2at}(2 \sin h\; at) = \frac{1}{at} \sin h\,(at)$$ **Hence proved.**

Again $M_0(t) = (1/at) \sin h\,(at)$

$$= \frac{1}{at}\left[(at) + \frac{(at)^3}{3!} + \frac{(at)^5}{5!} + \ldots\right],$$

$$\because \quad \sin h = x + \frac{x^3}{3\,!} + \frac{x^5}{5\,!} + \ldots\ldots$$

$$\text{or} \quad M_0(t) = 1 + \frac{a^2 t^2}{3\,!} + \frac{a^4 t^4}{5\,!} + \ldots + \frac{a^{2r} t^{2r}}{(2r+1)\,!} + \ldots, \qquad ..(i)$$

$\therefore \mu_{2r} =$ coefficient of $\frac{t^{2r}}{(2r)\,!}$ in the expansion of $M_0(t)$

$$= \frac{a^{2r}}{(2r+1)\,!} \times (2r)\,!, \text{ from (i)} \qquad \textbf{(Note)}$$

$$= a^{2r}/(2r+1)$$

Hence proved.

Q. 20. Define a Poisson distribution. Obtain the recurrence formula : $\mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda}$ for this distribution and hence, obtain the first four moments.

Ans. Given that $\mu_r = \sum_{j=0}^{\infty}\left[\frac{e^{-\lambda}\lambda^j(j-\lambda)^r}{j\,!}\right]$ ..(i)

$$\therefore \frac{d\mu_r}{d\lambda} = \sum_{j=0}^{\infty}\left[\frac{1}{(j)\,!}\frac{d}{d\lambda}\{e^{-\lambda}\lambda^j(j-\lambda)^r\}\right]$$

$$= \sum_{j=0}^{\infty}\left[\frac{1}{(j)\,!}\{-e^{-\lambda}\lambda^j(j-\lambda)^r + e^{-\lambda}j\lambda^{j-1}(j-\lambda)^r\}\right]$$

$$+ e^{\lambda}\lambda^j r(j-\lambda)^{r-1}(-1)3]$$

$$= \sum_{j=0}^{\infty}\frac{1}{(j)\,!}[e^{-\lambda}\lambda^{j-1}(j-\lambda)^r\{-\lambda+j\} - re^{-\lambda}\lambda^j(j-\lambda)^{r-1}]$$

$$\text{or} \quad \lambda\frac{d\mu_r}{d\lambda} = \sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^j(j-\lambda)^{r+1}}{(j)\,!} - \lambda r\sum_{j=0}^{\infty}\frac{e^{-\lambda}\lambda^j(j-\lambda)^{r-1}}{(j)\,!},$$

multiplying each term on both sides by λ

$$= \mu_{r+1} - r\lambda\mu_{r-1}, \text{ from (i)}$$

$$\text{or} \quad \mu_{r+1} = r\lambda\mu_{r-1} + \lambda\frac{d\mu_r}{d\lambda} \qquad ..(i)$$

Hence proved

Also we know that for Poisson distribution, $\mu_1 = 0$.
Putting $r = 1, 2, 3$ in (i) we obtain

$$\mu_2 = 1 \,.\, \lambda \,\mu_0 + \lambda \frac{d\mu_1}{d\lambda} = 1 \,.\, \lambda \,.\, 1 + 0 = \lambda;$$

$$\mu_2 = 2\lambda \,\mu_2 + \lambda \frac{d\mu_2}{d\lambda} = 2 \,.\, \lambda \,.\, 0 + \lambda \frac{d\lambda}{d\lambda} = \lambda;$$

$$\mu_4 = 3\lambda \,\mu_2 + \lambda \frac{d\mu_3}{d\lambda} = 3\lambda \,\lambda + \lambda \frac{d\lambda}{d\lambda} = 3\lambda^2 + \lambda$$

●

Q. 21. Show that in a Poisson distribution with unit mean, mean deviation about mean is (2/e) times the standard deviation.

Ans. We know the probability of x successes in Poisson distribution $= e^{-m}\, m^x/(x\,!)$

Here we are given mean = 1 *i.e.*, $m = 1$

$\therefore$ Standard deviation $= \sqrt{m} = \sqrt{(1)} = 1$ ..(i)

And the probability of x successess $= e^{-1}/(x\,!)$

Now mean deviation about the mean

$$= \sum_{x=0}^{\infty} |\, x - m \,|\, P(x), \text{ where } x = 1 \qquad \textbf{(Note)}$$

$$= \sum_{x=0}^{\infty} |\, x - 1 \,| \frac{e^{-1}}{x\,!}$$

$$\because \quad P(x) = \frac{e^{-1}}{x\,!}$$

$$= \frac{1}{e} \sum_{x=0}^{\infty} \frac{|\, x - 1 \,|}{x\,!} = \frac{1}{e}\left[1 + \frac{1}{2\,!} + \frac{2}{3\,!} + \frac{3}{4\,!} + \ldots\right] \qquad \text{..(ii)}$$

General term of this series $= \dfrac{n}{(n+1)\,!} = \dfrac{(n+1) - 1}{(n+1)\,!}$

$$= \frac{1}{(n)} - \frac{1}{(n+1)\,!}$$

Putting $n = 1, 2, 3, \ldots$ etc., we get from (ii) the mean deviation about mean

$$= \frac{1}{e}\left[1 + \left(\frac{1}{1\,!} - \frac{1}{2\,!}\right) + \left(\frac{1}{2\,!} - \frac{1}{3\,!}\right) + \right]$$

$= (2/e) = (2/e) \times 1$ **(Note)**

$= (2/e) \times$ standard deviation from (i).

Hence proved.

●

UNIT

2

Your Syllabus

Distributions of function of random variables : Distribution of sum, product and quotient of two Variable. Reproductive property of standard distribution. χ^2 (chi-square), t and F distributions (Central cases only) and their limiting forms. Bivariate normal distribution and its properties.

Compulsory Questions

(1) What do you understand by chi-square distribution. What are its properties ?

Ans. ψ^2distribution is a continuous distribution whose probability density function is given by

$P(\psi^2) = y_0 (\psi^2)^{\frac{1}{2}(v-2)} e^{-\frac{1}{2}\psi^2}$, where v = degrees of freedom $= n - 1$

and y_0 = constant depending on the degrees of freedom.

Properties of ψ^2 – Distribution

1. Chi-square curve is always positively skewed.

2. The mean of ψ^2 distribution is the number of degrees of freedom.

3. The standard deviation of ψ^2 – distribution $= \sqrt{2v}$, where v is the degrees of freedom.

4. Chi-square values increase with the increase in degrees of freedom.

5. The value of ψ^2 lies between zero and infinity, *i.e.*, $0 \leq \psi^2 < \infty$.

6. The sum of two ψ^2-distribution is again a ψ^2 distribution, *i.e.*, if ψ_1^2 and ψ_2^2 are two independent and they have a ψ^2 distribution with n_1 and n_2 degrees of freedom respectively, then $(\psi_1^2 + \psi_1^2)$ is also a ψ^2-distribution with $(n_1 + n_2)$ degrees of freedom.

7. Its shape depends on the degree of freedom but it is not a symmetrical distribution.

(2) What are the uses of Chi-square tests ?

Ans. The important uses of ψ^2 test are :

1. **Test of Goodness of Fit :** If the two curves, *viz*, (i) Observed frequency curve and (ii) the expected frequency curve, are drawn, then the Chi-square statistic may be used to determine whether the two curves so drawn are fitted good or not. Thus, the term goodness of fit is used to test the concordance of the fitness of these two curves. Under this test there is only one variable, so the degrees of freedom is $\nu = n - 1$.

2. **Test of Independence of Attribute :** The Chi-square test is used to see that the principles of classification of attributes are independent. In this test the attributes are classified into a two way table or a contingency table as the case may be. The observed frequency in each cell (square) is known as **Cell frequency.** The total frequency in each row or column of the two way contingency table is known as **Marginal frequency.**

The degrees of freedom is $\nu\,(R - 1)\,(C - 1)$,

where R = number of rows, C = number of columns in the two way contingency table. This test discloses whether there is any association or relationship between two or more attributes.

3. Test of Homogeneity or a Test for a Specified Standard deviation : The Chi-Square test may be used to test the homogeneity of the attributes in respect of a particular characteristic or it may also be used to test the population variance. In the case of a specified standard deviation the test statistic is given by $\psi^2 = (n - 1)\,s^2/\sigma_0^2$, where s^2 = sample variance and σ_0^2 is the hypothesized value of population variance. ●

(3) What are the conditions applying the chi-square tests ?

Ans. 1. Each of the observations making up the sample for this test should be independent of each other.

2. The expected frequency of any item or cell should not be less than 5. If it is less than 5, then frequencies taking from the adjacent items or cells are pooled together in order to make it 5 or more than 5.

3. The total number of observations used in this test must be large *i.e.*, $n \geq 30$.

4. This test is used only for drawing inferences by testing hypothesis. It cannot be used for estimation of parameter or any other value.

5. It is wholly dependent on the degrees of freedom.

6. The frequencies used in ψ^2-test should be absolute and not relative in terms.

7. The observations collected for ψ^2-test should be on random basis of sampling. ●

(4) Explain Chi-square distribution of variance.

Ans. Let σ^2, s^2 respectively be the population variance and sample variance. The sampling distribution $(n - 1)\,s^2/\sigma^2$ has a chi-square (ψ^2) distribution with $n - 1$ degrees of freedom.

It is a very useful distribution in making inferences about the population variance σ^2 by using sample variance s^2. It is used in making interval estimate of the population variance, which is given by the result,

$$\frac{(n-1)\,s^2}{\psi^2_{\alpha/2}} \le \alpha < \frac{(n-1)\,s^2}{\psi^2_{(1-\alpha/2)}},$$

where α is the level of significance and ψ^2_{α} = value for the chi-square distribution giving an area to the right of stated ψ^2_{α} and $(1-\alpha)$ confidence coefficient.

➲ (5) Describe the chi-square test of significance and state the various uses to which it can be put.

Or

What is χ^2 test ? Point out its uses.

Ans. Chi-square is a measure which evaluates extent to which a set of the observe frequencies of a sample deviates from the corresponding set of theoretical (expected) frequencies of the sample. It is a measure of the aggregate discrepancy between the actual frequencies and the theoretical frequencies in a sample.

Chi-square is used as a test statistics in testing a hypothesis that provides a set of theoretical frequencies with which observed frequencies are compared. The measure of chi-square enables us to find out the degree of discrepancy between observed frequencies and theoretical frequencies and thus to determine whether the discrepencies so obtained is due to error of sample or due to chance.

In other words, a test statistics which measures the discrepancy between observed or actual frequencies *i.e.,* $O_1, O_2, \ldots, O_n$ and their corresponding expected frequencies, *i.e.,* $E_1, E_2, E_3, \ldots, E_n$ is called the chi-square (x_2) statistics which can be obtained by using the formula

$$X^2 = \Sigma\frac{(O_i - E_i)^2}{E_i} = \left[\frac{(O_1 - E_1)^2}{N} + \frac{(O_2 - E_2)^2}{E_2} + \ldots + \frac{(O_n - E_n)^2}{E_n}\right]$$

i.e., the difference between the observed and expected frequencies, is squared divided by the expected frequencies, the chi-square (χ^2) statistic is the sum of the quotient of all the cells or categories.

Degree of Freedom (d.o.f.) : The degrees of freedom play a very vital role in χ^2 test of a hypothesis. The chi-square test depends on the number of degree of freedom. The data that are given in the form of a series of variables in a row or column, or the number of frequencies put in cells in a contingency table, which can be calculated independently, are knows as degree of freedom (ν).

In case of data that are given in the form of a series of variables in a row or column, the degree of freedom will be the number of items in the series less one *i.e.,* ($\nu = n - 1$). Similarly, in case of the number of frequencies put in cells in a contingency table, the degree of freedom will be equal to product of the number of rows less one and the number of columns less one. $V = (R - 1)(C - 1)$.

For a given set of observed frequencies, (say n) which are subject to k independent restrictions (constraints) then Degree of Freedom (d.o.f.) is $V = n - k$.

Application of Chi-square Test

The application of χ^2 have been listed below :

(I) Chi-square test of goodness of fit,

(II) χ^2-test for independence of attributes,

(III) To test if the population has a specified value of the variance σ^2,

(IV) Test of homogeneity.

➲ **(6) What do you mean by *F*-statistic ? Discuss the derivation of Snedecor's *F*-distribution. What are the constants of *F*-distribution?**

Ans. If X and Y are two independent chi-square variates with ν_1 and ν_2 d.f. respectively, then F-statistic is defined by

$$F = \frac{X/\nu_1}{Y/\nu_2}$$

In other words, F is defined as the ratio of two independent chi-square variates divided by their respective degrees of freedom and it follows Snedecor's F-distribution with (ν_1, ν_2) d.f. with probability function given by

$$f(F) = \frac{\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \cdot \frac{F^{\frac{\nu_1}{2} - 1}}{\left[1 + \frac{\nu_1}{\nu_2}F\right]^{(\nu_1 + \nu_2)/2}}, \quad 0 \le F < \infty$$

➲ **(7) What do you mean by Random Variable ? Give the statement of some of the theorems on Random Variables.**

Ans. Random Variable. By a random variable (r, v) we mean a real number X connected with the outcome of a random experiment E. For example, if E consists of two tosses of a coin, we may consider the random variable which is the number of heads (0, 1 or 2).

Outcome :	HH	HT	TH	TT
Value of X :	2	1	1	0

Thus to each outcome ω, there corresponds a real number $X(\omega)$. Since the points of the sample space S correspond to outcomes, this means that a real number, which we denote by $X(\omega)$, is defined for each

$\omega \in S$. From this point of view we define random variable to be a real functoin on S as follows :

"Let S be the sample space associated with a given random experiment. A real-valued function defined on S and taking values in R (− ∞, ∞) is called a one-dimensional random variable. If the function values are ordered pairs of real numbers (i.e., vectors in two-space) the function is said to be a two-dimensional random variable. More generally, an n-dimensional random variable is simply a function whose domain is S and whose range is a collection of n-tuples of real numbers (vectors in n-space)."

For a mathematical definition of the random variable, let us consider the probability space, the triplet (S, B, P), where S is the sample space, viz., space of outcomes, B is the σ-field of subsets in S and P is a probability function on B.

Def. A random variable (r, v) is a function $X(\omega)$ with doman S and range $(-\infty, \infty)$ such that for every real number a, the event $[\omega : X(\omega) \le a] \in B$.

Notes : 1. The refinement above is the same as saying that the function $X(\omega)$ is measurable real function on (S, B).

2. We shall need to make probability statements about a random variable X such as $P(X \le a)$. For the simple example given above we should write $P(X \le 1) = P$ (HH, HT, TH) = 3/4. That is, $P(X \le a)$ is simply the probabilityof the set of outcomes ω for which $X(\omega) \le a$ or

$$P(X \le a) = P\{\omega : X(\omega) \le a\}.$$

Since P is a measure on (S, B) *i.e.*, P is defined on subsets of B, the above probability will be defined *only* if $\{\omega : X(\omega) \le a\} \in B$, which implies that $X(\omega)$ is a mesurable function on (S, B).

Examples : 1. If a coin is tossed, then

$$S = \{\omega_1, \omega_2\} \quad \text{where } \omega_1 = H, \omega_2 = T$$

$$X(\omega) = \begin{cases} 1, \text{if } \omega = H \\ 0, \text{if } \omega = T \end{cases}$$

$X(\omega)$ is a Bernoulli random variable. Here $X(\omega)$ takes only two values. A random variable which takes only a finite number of values is called *single.*

2. An experiment consists of rolling a die and reading the number of points on the upturned face. The most natural random variable X to consider is

$$X(\omega) = \omega ; \omega = 1, 2, \ldots 6.$$

If we are interested in whether the number of points is even or odd, we consider a random variable Y defined as follows :

$$Y(\omega) = \begin{cases} 0, \text{if } \omega \textit{ is even} \\ 1, \text{if, } \omega \textit{ is odd} \end{cases}$$

3. If a dart is thrown at a circular target, the sample space S is the set of all points w on the target. By imagining a coordinate system placed

on the target with the origin at the centre, we can assign various random variables to this experiment. A natural one is the two dimensional random variable which assigns to the point ω, its rectangular coordinates (x, y). Another is that which assigns to each ω only one of the coordinates x or y (for cartesian system), r or θ (for polar system). The event E, "that the dart will land in the first quadrant" can be described by a random variable which assigns to each point w its polar coordinate θ so that $X(\omega) = \theta$ and then $E = \left\{\omega : 0 \le X(\omega) \le \pi/2\right\}$.

Theorem 1. A function $X(\omega)$ from S to $R(-\infty, \infty)$ is a random variable if and only if

$$\{\omega : X(\omega) < a\} \in \mathbf{B}.$$

Theorem 2. If X_1 and X_2 are random variables and C is a constant then $CX_1, X_1 + X_2, X_1 X_2$ are also random variables.

Theorem 3. If $\{X_n(\omega), n \ge 1\}$ are random variables then $\sup\limits_{n} X_n(\omega)$, $\inf\limits_{n} X_n(\omega)$, $\limsup\limits_{n \to \infty} X_n(\omega)$ and $\liminf\limits_{n \to \infty} X_n(\omega)$ are all random varaibles, whenever they are finite for all ω.

Theorem 4. If X is a random variable then

(i) $\dfrac{1}{X}$ where $\left(\dfrac{1}{X}\right)(\omega) = \infty$ if $X(\omega) = 0$

(ii) $X+(\omega) = \max[0, X(\omega)]$

(iii) $X-(\omega) = -\min[0, X(\omega)]$

(iv) $|X|$

are random variables.

Theorem 5. If X_1 and X_2 are random variables then

(i) $\max[X_1, X_2]$ and (ii) $\min[X_1, X_2]$ are also random variables.

Theorem 6. If X is a r.v. and $f(.)$ is a continuous function, then $f(X)$ is a r.v.

Theorem 7. If X is a r.v. and $f(.)$ is an increasing function, then $f(X)$ is a r.v. ●

➲ (8) What do you mean by distribution function ? Prove that if F is the distribution function of one dimensional random variable X, then

$$f(-\infty) = \lim_{x \to -\infty} f(x) = 0$$

and

$$f(\infty) = \lim_{x \to \infty} f(x) = 1.$$

Ans. Distribution Function. Let X be a r.v. on (S, B, P). Then the function :

$$F_X(x) = P(X \le x) = P\left\{\omega : X(\omega) \le x\right\} - \infty < x < \infty \qquad \ldots(1)$$

is called the distribution function (d.f.) of X.

For the purpose of clarity we may write $F(x)$ instead of $F_X(x)$.

Properties of Distribution Function.

Property 1. If F is the d.f. of the r.v. X and if $a < b$ then

$$P(a < X \le b) = F(b) - F(a)$$

Proof. The events '$a < X \le b$' and '$X \le a$' are disjoint and their union is the event '$X \le b$'. Hence by addition theorem of probability

$$P(a < X \le b) + P(X \le a) = P(X \le b)$$

$$\Rightarrow P(a < X \le b) = P(X \le b) - P(X \le a) = F(b) - F(a) \quad ...(2)$$

Cor. 1.

$$P(a < X \le b) = P\{(X = a) \cup (a < X \le b)\}$$
$$= P(X = a) + P(a < X \le b)$$

(using additive property of P)

$$= P(X = a) + [F(b) - F(a)] \quad ...(3)$$

Similarly, we get

$$P(a < X < b) = P(a < X \le b) - P(X = b)$$
$$= F(b) - F(a) - P(X = b) \quad ...(4)$$

$$P(a \le X < b) = P(a<X<b) + P(X = a)$$
$$= F(b) - F(a) - P(X = b) + P(X = a) \quad ...(5)$$

Note : When $P(X = a) = 0$ and $P(X = b) = 0$, all four events $a \le X \le b, a < X < b, a \le X < b$ and $a < X \le b$ have the same probability $F(b) - F(a)$.

Property 2. If F is the d.f. of one-dimensional r.v. X, then

(i) $0 \le F(x) \le 1$,

(ii) $F(x) \le F(y)$ if $x < y$, *i.e. all distribution functions are monotonically non-decreasing and lie between 0 and 1.*

Proof. Using the axioms of certainity and non-negativity for the probability function **P**, part (i) follows triviality from the definition of $F(x)$.

For part (ii), we have for $x<y$,

$$F(y) - F(x) = P(x < X \le y) \ge 0 \quad \text{(Property 1)}$$

$$\Rightarrow \quad F(y) \ge F(x)$$

$$\Rightarrow \quad F(x) \le F(y) \text{ when } x < y \quad ...(6)$$

Property 3. If F is d.f. of one-dimensional r.v. X, then

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$

and $$F(\infty) = \lim_{x \to \infty} F(x) = 1.$$

Proof. Let us express the whole sample space S as a countable union of disjoint events as follows :

$$S = [\bigcup_{n=1}^{\infty} (-n < X \le -n + 1)] \cup [\bigcup_{n=0}^{\infty} (n < X \le n +)]$$

$$\Rightarrow P(S) = \sum_{n=1}^{\infty} P(-n < X \le -n + 1) + \sum_{n=0}^{\infty} P(n < X \le n + 1)$$

($\because$ P is additive)

$$\Rightarrow 1 = \lim_{a \to \infty} \sum_{n=1}^{a} [F(-n+1) - F(-n)$$

$$+ \lim_{b \to \infty} \sum_{n=0}^{b} [F(n+1) - F(n)]$$

$$= \lim_{a \to \infty} [F(0) - F(-a)] + \lim_{b \to \infty} [F(b+1) - F(0)]$$

$$= [F(0) - F(-\infty)] + [F(\infty) - F(0)]$$

$$\therefore \quad 1 = F(\infty) - F(-\infty) \qquad \text{...(7)}$$

Since $-\infty < \infty, F(-\infty) \leq F(\infty)$. Also

$$F(-\infty) \geq 0 \text{ and } F(\infty) \leq 1 \qquad \text{(Property 2)}$$

$$\therefore \quad 0 \leq F(-\infty) \leq F(\infty) \leq 1 \qquad \text{...(8)}$$

(7) and (8) give $F(-\infty) = 0$

and $F(\infty) = 1$. ●

➲ **(9) Write a short note on Bivariate Normal Distribution.**

Ans. Bivariate Normal Distribution : The bivariate normal distribution is a generalization of a normal distribution for a single variate. Let X and Y be two normally correlated variables with correlation coefficient ρ and $E(X) = \mu_1$, Var $(X) = \sigma_1^2$; $E(Y) = \mu_2$, Var $(Y) = \sigma_2^2$. In deriving the bivariate normal distribution we make the following three assumptions.

(i) The regression of Y on X is linear. Since the mean of each array is on the line of regression $Y = \rho(\sigma_2/\sigma_1)X$, the mean or expected value of Y is $\rho(\sigma_2/\sigma_1)X$, for different values of X.

(ii) The arrays are homoscedastic, i.e., variance in each array is same. The common variance of estimate of Y in each array is then given by $\sigma_2^2(1-\rho^2)$, ρ being the correlation coefficient between variables X and Y and is independent of X.

(iii) The distribution of Y in defferent arrays in normal. Suppose that one of the variates, say X, is distributed normally with mean 0 and standerd deviation σ_1 so that the probality that a random value of X will fall in the small interval dx is

$$g(x)\,dx = \frac{1}{\sigma_1\sqrt{(2\pi)}} \cdot \exp(-x^2/2\sigma_1^2)\,dx.$$

The probality that a value of Y, taken at random in an assigned vertical array will fall in the interval dy is

$$h(y|x)\,dy = \frac{1}{\sigma_2\sqrt{2\pi(1-\rho^2)}} \; exp.\left\{\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y - \rho x\frac{\sigma_2}{\sigma_1}\right)^2\right\}$$

The joint probability differential of X and Y is given by

$$dP(x,y) = g(x)\, h(y|x)\,dx\,dy$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}}\, e^{\frac{1}{2\sigma_1^2}x^2}\, e^{\left[-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y-\rho\frac{\sigma_2}{\sigma_1}x\right)^2\right]}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right)\right\}$$

Shifting the origin to (μ_1, μ_2), we get

$f_{XY}(x, y)$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \cdot e^{-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right\}}$$

$$(-\infty < x < \infty, -\infty < y < \infty)$$

where $\mu_1, \mu_2, \sigma_1 (>0), \sigma_2 (>0)$ and $\rho\ (-1 < \rho < 1)$ are the five parameters of the distribution.

This is the density function of a bivariate normal distribution. The variables X and Y are said to be normally correlated and the surface $z = f(x, y)$ is known as the normal correlation surface. ●

☛ (10) What do you understand by student-distribution ? What are the various properties of t-distribution ?

Ans. Theoretical work on *t*-distribution was done by W.S. Gosset (1876-1937) in the early 1900. The *t*-distribution is used when sample size is 30 or less and the population standard deviation is unknown.

The "*t*-statistic" is defined as :

$$t = \frac{\overline{X} - \mu}{S} \times \sqrt{n}$$

where $$S = \sqrt{\frac{\Sigma(X - \overline{X})^2}{n-1}}$$

The *t*-distribution has the following form

$$f(t) = C\left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}$$

where $$t = \frac{(\overline{X} - \mu)}{S}\sqrt{n}$$

C = a constant required to make the area under the curve equal to unity v = n − 1, the number of degrees of freedom.

Properties of *t*-Distribution

1. The variable *t*-distribution ranges from − ∞ to + ∞.

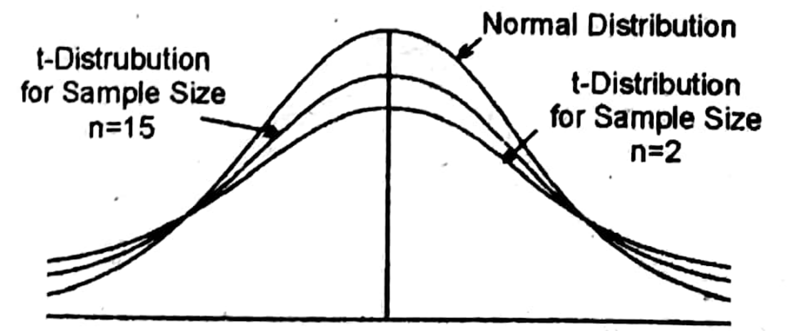
2. The constant C is actually a function of ν, so that for a particular value of ν, the distribution of $f(t)$ is completely specified. Thus $f(t)$ is a family of functions, one for each value of ν.

3. Like the standard normal distribution, the t-distribution is symmetrical and has a mean zero.

4. The variance of t-distribution is greater than one, but approaches one as the number of degrees of freedom and, therefore, the sample size becomes large. Thus the variance of the t-distribution approaches the variance of the standard normal distribution as the sample size increases. It can be demonstrated that from an infinite number of degrees of freedom ($\nu = \infty$), the t-distribution and normal distribution are exactly equal. Hence there is a widely practised rule of thumb that samples of size n 30 may be considered large and the standard normal distribution may appropriately be used as an approximation to t-distribution, where the latter is the theoretically correct functional form.

The following diagram compares one normal distribution with two t-distributions of different sample sizes :

The above diagram shows two important characteristics of t-distribution. First, a t-distribution is lower at the mean and higher at the tails than a normal distribution. Second, the t-distribution has proportionately greater area in its tails than the normal distribution. ●



Normal Distribution, t-Distribution for Sample Size n=15, and t-Distribution for Sample Size n=2

(11) Discuss the Fisher's f-Test for testing the equality of two sample variance.

Ans. The object of the F-test is to find out whether the two independent estimates of population variance differ significantly, or whether the two samples may be regarded as drawn from the normal populations having the same variance. For carrying out the test of significance, we calculate the ratio F. F is defined as :

$$F = \frac{S_1^2}{S_2^2}, \text{ where } S_1^2 = \frac{(X_1 - \bar{X}_1)^2}{n_1 - 1}$$

and

$$S_2^2 = \frac{\Sigma\,(X_2 - \bar{X}_2)^2}{n_2 - 1}$$

It should be noted that S_1^2 is always the larger estimate of variance.

i.e., $S_1^2 > S_2^2$

$$F = \frac{\text{Larger estimate of variance}}{\text{Smaller estimate of variance}}$$

$v_1 = n_1 - 1$ and $v_2 = n_2 - 1$

v_1 = degrees of freedom for sample having larger variance.

v_2 = degrees of freedom for sample having smaller variance.

The calculated value of F is compared with the table value for v_1 and v_2 at 5% or 1% level of significance. If calculated value of F is greater than the table value then the F ratio is considered significant and the null hypothesis is rejected. On the other hand, if the calculated value of F is less than the table value the null hypothesis is accepted and it is inferred that both the samples have come from the population having same variance.

Since F test is based on the ratio of two variance's it is also known as the Variance Ratio Test. The ratio of two variances follows a distribution called the F distribution named after the famous statistician R.A. Fisher.

Assumptions in *F*-Test The F-test is based on the following assumptions :

1. **Normality.** *i.e.*, the values in each group are normally distributed :
2. **Homogeneity.** *i.e.*, the variance within each group should be equal for all groups ($\sigma_1^2 = \sigma_2^2 = ... = \sigma_c^2$) This assumption is needed in order to combine or pool the variances within the groups into a single 'within groups' source of variation.
3. **Independence of error.** It states that the error (variation of each value around its own group mean) should be independent for each value.

(12) Write Conditions for the application of χ^2 test.

Ans. Before the application of χ^2-test, the following conditions must be satisfied :

(i) N (*i.e.*, the total frequency) must be sufficiently large; not less than 50 however small the number of cells may be.

(ii) The smallest cell frequency should be 10 or more. If the frequencies are less than 10 in two or more cells, then these should be combined into a single cell having frequency 10 or more.

(iii) All the individuals in the sample must be independent.

(iv) The constraints imposed on the cell frequencies must be linear,

(v) The number of cells must be between 5 and 20. The number of cells less than 5 can be used only if the cell frequencies are not small.

(13) The following table gives the number of aircraft accidents that occurred during the various days of the week :

Days	Sun.	Mon.	Tue.	Wed.	Thu.	Fri.	Sat
No. of accidents	14	16	8	12	11	9	14

Using x^2-test find whether the accidents are uniformaly distributed over the week. Given that x^2 for 6 degreesof freedom at 5% level = 12·59.

Ans. The total number of accidents (observed) in the week

$$= 14 + 16 + 8 + 12 + 11 + 9 + 14 = 84.$$

Had the accidents been uniformly distributed over the week, then the expected number of accidents every day of the week would have been (84/7) *i.e.*, 12.

$$\therefore \chi^2 = \frac{(14-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(8-12)^2}{12} + \frac{(12-12)^2}{12} + \frac{(11-12)^2}{12} + \frac{(9-12)^2}{12} + \frac{(14-12)^2}{12}$$

$$= \frac{1}{12}\left[4 + 16 + 16 + 0 + 1 + 9 + 4\right] = \frac{50}{12} = 4.17$$

Also here the number of degrees of freedom $= \nu = n - 1 = 7 - 1 = 6$. And the given value of χ^2 for six degrees of freedom at 5% level = 12.59.

Thus the calculated value of χ^2 viz. is less than this value and so the hypothesis holds good *i.e.*, the accidents are uniformly distributed over the week.

●

➲ **(14) 200 digits were chosen at random from a table. The frequencies of the digits were :**

Digit :	0	1	2	3	4	5	6	7	8	9
Frequency :	18	19	23	21	16	25	22	20	21	15

Use the χ^2-test to access hypothesis that the digits were distributed in equalnumbers in the table from which these numbers were taken.

Ans. Here the total frequency

= 18 + 19 + 23 + 21 + 16 + 15 + 22 + 20 + 21 + 15
= 200

$\therefore$ The theoretical frequencies for each digit $= \frac{200}{10} = 20$ *i.e.*, f_e for each digit is 20.

$$\text{Then } \chi^2 = \Sigma\left[\frac{(f_o - f_e)^2}{f_e}\right] = \frac{(18-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(23-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(16-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(22-20)^2}{20}$$

$$+ \frac{(20-20)^2}{20} + \frac{(21-20)^2}{20} + \frac{(15-20)^2}{20}$$

$$= \frac{1}{20}[4+1+9+1+16+25+4+0+1+25] = \frac{86}{20} = 4.30$$

Also the number of degrees of freedom

$$= v = n - 1 = 10 - 1 = 9.$$

From the χ^2 table we find that 5% value of the chi-square for nine degrees of freedom = 16.919.

The calculated value of χ^2 viz. 4.30 is much less than this value and so this calculated value is not significance and so the hypothesis appears to be true.

(15) Write about the Student's t-distribution.

Ans. **Student's t-distribution.**

Student's t-distribution was founded by W.S. Gusset who published it under the pen name 'Student'.

Student's t-distribution with $(n-1)$ degree of freedom is defined as the distribution of the statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

where $\bar{x}$ is the mean of the sample of size n drawn from a normal population of mean μ and s^2 is an unbiased estimate of population variance

i.e.,
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

(16) Find student's t for the following variate in a sample of eight taking the mean of the universe to be zero.

– 4, – 2, – 2, 0, 2, 2, 3, 3.

Ans. Let the assumed mean $a = 2$

x	$d_i = x - a$	dt^2
– 4	– 6	36
– 2	– 4	16
– 2	– 4	16
0	– 2	4
2	0	0
2	0	0
3	1	1
3	1	1
Total	$\Sigma d_i = -14$	$\Sigma d_i^2 = 74$

$$\therefore\ \bar{x} = a + \frac{\Sigma d_i}{n} = 2 + \frac{(-14)}{8} = 2 - \frac{7}{4}$$

$$= \frac{1}{4} = 0.25$$

$$s^2 = \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right] = \frac{1}{8-1}\left[74 - \frac{(-14)^2}{8}\right]$$

$$= \frac{1}{7}\left[74 - \frac{49}{2}\right] = \frac{1}{7}\left[\frac{99}{2}\right] = \frac{99}{14} = 7.071$$

$$\therefore\ s = \sqrt{(7.071)} = 2.659$$

Now $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$, where $\bar{x} = 0.25$, $\mu = 0$ (given),

$s = 2.659, n = 8$

$$\therefore\ t = \frac{0.25 - 0}{(2.659)/\sqrt{8}} = \frac{(0.25)\sqrt{8}}{2.659}$$

$$= \frac{(0.25)(2.828)}{2.659} = \frac{0.707}{2.659}$$

$$= 0.2658 = 0.27$$

➲ (17) Ten individuals are chosen at random from a normal population and their heights are found to be
63, 63, 66, 67, 68, 69, 70, 70, 71, 71 inches.
Test if the sample belongs to the population whose mean height is 66 inches. Given *t* for –9 d.f. at 5% level of significance is 2.262.

Ans. Let the assumed mean '*a*' = 68. Then.

x	$d = x - a$	d^2	x	$d = x - a$	d^2
63	– 5	25	70	2	4
63	– 5	25	70	2	4
66	– 2	4	71	3	9
67	– 1	1	71	3	9
68	0	0	**Total**	– 2	82
69	1	1			

$$\therefore\ \bar{x} = a + \frac{\Sigma d_i}{n} = 68 + \frac{(-2)}{10} = 68 - 0\cdot2 = 67\cdot8$$

$$s^2 = \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right] = \frac{1}{10-1}\left[82 - \frac{(-2)^2}{10}\right]$$

or $s^2 = \frac{1}{9}[82 - 0\cdot4] = \frac{1}{9}[81\cdot6] = 9\cdot067$

or $s = 3{\cdot}011$

Now $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}} = \dfrac{67{\cdot}8 - 66{\cdot}0}{(3{\cdot}011)/\sqrt{(10)}}$

$$= \frac{(1{\cdot}8)\cdot\sqrt{(10)}}{3{\cdot}011} = \frac{(1{\cdot}8)\,(3{\cdot}162)}{3{\cdot}011}$$

$$= \frac{5{\cdot}692}{3{\cdot}011} = 1{\cdot}890$$

Now the calculated value of t viz. $1{\cdot}89$ is less than the given value (or tbale value) $2{\cdot}262$. The difference between $\bar{x}$ and μ is not significant and so we may conclude that the sample has been drawn from the population of mean height 66 inches.

☛ (18) Ten individuals are chosedn at random from a population and their heights are found to be in inches :
63, 63, 66, 67, 68, 69, 71, 70, 71, 70
In the light of these data, discuss the suggestion that the mean height in the population in 66 inches having given tbat for $t = 1{\cdot}8$, $P = 0{\cdot}947$ and for $t = 1{\cdot}9$, $P = 0{\cdot}955$, where P is the area to the leftof the ordinate at t.

Ans. We know $t = \dfrac{\bar{x} - \mu}{s/\sqrt{(n)}}$, ..(i)

$\bar{x}$ and s are calculated as given below :
Let the assumed mean $A = 67$

x	$d = x - A = x - 67$	d^2
63	– 4	16
63	– 4	16
66	– 1	1
67	0	0
68	1	1
69	2	4
71	4	16
70	3	9
71	4	16
70	3	9
Total	$\Sigma d = 8$	$\Sigma d^2 = 88$

$$\therefore\ \bar{x} = A + \frac{\Sigma d}{n} = 67 + \frac{8}{10} = 67.8 \text{ inches}$$

$$s^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right] = \frac{1}{10-1}\left[88 - \frac{(8)^2}{10}\right]$$

$$= \frac{1}{9}[88 - 6.4 = \frac{1}{9}[\quad = 9.066$$

or $s = \sqrt{(9.066)} = 3.011$

$\therefore$ From (i), $t = \dfrac{\bar{x}-\mu}{s/\sqrt{n}} = \dfrac{\bar{x}-\mu}{s}\sqrt{n} = \dfrac{67.8-66}{(3.011)}.\sqrt{(10)}$

$$= \frac{(1.8)\,(3.162)}{3.011} = \frac{5.692}{3.011} = 1.89$$

Now difference for 0.1 = 0.955 − 0.947 = 0.008

$\therefore$ Difference for 1.89 − 1.8 *i.e.*, $0{\cdot}99 = \dfrac{0.008}{0.1} \times 0.09 = 0.0072$

Hence for $t = 1.89$, $P = 0.947 + 0.0072 = 0.9542 = 0.954 = P_s$

$\therefore P_f = 2(1 - P_s) = 2(1 - 0.954) = 2(0.046) = 0.092 > 0.05$

We therefore conclude that the difference is not significant at 5% level of significance and the test does not provide any evidence against the population mean being 66 inches. ●

➲ (19) Write M.G.F. of Chi-Square Distribugion

Ans. **M.G.F. of CHI-Square Distribution**

Let $X \sim \chi^2(n)$, then

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x)\,dx = \frac{1}{2^{n/2}\,\Gamma(n/2)}$$

$$\int_0^\infty e^{tx}.e^{-x/2}x^{(n/2)-1}\,dx$$

$$= \frac{1}{2^{n/2}\,\Gamma(n/2)}\int_0^\infty \exp\left\{\left[-\left(\frac{1-2t}{2}\right)x\right].x^{(n/2)-1}\,dx\right.$$

$$= \frac{1}{2^{n/2}\,\Gamma(n/2)}\,\frac{\Gamma(n/2)}{[(1-2t)/2]^{n/2}} \quad \text{[Using Gamma Integral]}$$

$$= (1-2t)^{-n/2},\ |\,2t\,| < 1$$

which is the required *m.g.f.* of a χ^2-variate with *n.d.f.* ●

➲ (20) Additive or Re-productive Property of Noncentral Chi-Square Distribution : If Y_i, (i = 1, 2,...., k), are independent non-central χ^2-variates with n_i *d.f.* and non-centrality element λ_i then $\sum_{i=1}^{k} \lambda_i$.

Ans. We have

$$M_{Y_i}(t) = (1-2t)^{-n_i/2}\exp[2t\lambda_i/(1-2t)], \qquad (i = 1, 2, \ldots\ldots, k)$$

$$\therefore \quad M_{\Sigma Y_i}(t) \prod_{i=1}^{k} M_{Y_i}(t) = (1-2t)^{-\Sigma n_i/2}$$

$$\exp \quad [2t \sum_i \lambda_i/(1-2t)]$$

which is the m.g.f. of a non central χ^2-variate with $\sum_i n_i$ *d.f.* and non-centrality paramter $\lambda = \sum_i \lambda_i$. Hence by uniqueness theorem of m.g.f.'s

$$\sum_{i=1}^{k} Y_i \sim \chi'^2_{\Sigma n_i} (\sum_i \lambda_i)$$

Long Answer Type Questions

Q. 1. If χ_1^2 and χ_2^2 are two independent Chi-square variates with n_1 and n_2 degrees of freedom respectively, then show that $\left(\chi_1^2 + \chi_2^2\right)$ is also a Chi-square variate with $(n_1 + n_2)$ degrees of freedom.

Ans. If χ_1^2 and χ_2^2 be two independent χ^2 variates with n_1 and n_2 degrees of freedom, then $({\chi_1}^2 + {\chi_2}^2)$ is a χ^2-variate with $(n_1 + n_2)$ degrees of freedom.

Proof : m.g.f. of $({\chi_1}^2 + {\chi_2}^2)$

$$= (\text{m. g. f. of } {\chi_1}^2).(\text{m. g. f. of } {\chi_2}^2)$$

$$= (1-2t)^{-n_1/2}.(1-2t)^{-n_2/2}, \text{ from 4 result (v)}$$

$$= (1-2t)^{-\frac{1}{2}(n_1+n_2)}$$

Hence $({\chi_1}^2 + {\chi_2}^2)$ is a χ^2-variate with $(n_1 + n_2)$ degrees of freedom.

Q. 2. Define *F*-statistic and give the properties of *F*-distribution.

Ans. Let $x_1, x_2, ...x_{n_1}$ and $y_1, y_2, ...y_{n_2}$ be the values of two independent random samples drawn from two normal populations σ^2 having equal variances.

Let $\overline{x_1}$ and $\overline{x_2}$ be the sample means and $s_1^2 = \frac{1}{n_1-1}\sum_{1}^{n_1}(x_i - \bar{x})^2$, $s_2^2 = \frac{1}{n_2-1}\sum_{1}^{n_r}(y_i - \bar{y})^2$ be the sample variances.

Then we define *F* by the relation

$$F = \frac{s_1^2}{s_2^2}$$

$$(s_1^2 > s_2^2)$$

This gives F-distribution (also known as variance ratio distribution) with $\gamma_1 = n_1 - 1$ and $\gamma_2 = n - 2$ degrees of freedom. The larger of the variances is placed in the numerator.

(2) Properties I. The F-distribution curve lies entirely in the first quadrant and is unimodal.

II. The F-distribution is independent of the population variance σ^2 and depends on γ_1 and γ_2 only.

●

Q. 3. Define χ^2-distribution. Find its moment generating function andhence show that its variance is two times the mean.

Ans.

Introduction

The square of a standard normal variate is known as a chi-square variate (pronounced as Ki-Sky without S) with 1 degree of freedom (d.f.,)

Thus if $X \sim N(\mu, \sigma^2)$,

then $\quad Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$ and

$$Z^2 = \left(\frac{X - \mu}{\sigma}\right)^2 \text{ is a chi-square variate with 1 } d.f.$$

In general if X_i, $(i = 1, 2, \ldots\ldots, n)$ are n independent normal variates withmeans μ_i and variances σ_i^2, $(i = 1, 2, \ldots\ldots, n)$, then

$$\chi^2 = \sum_{x=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2, \text{ is a chi-square variate with } n\ d.f.$$

M.G.F., of Chi-Square Distribution

Let $X \sim \chi^{2(n)}$, then

$$M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} f(x)\, dx = 1 \quad 2^{n/2}\Gamma(n/2)$$

$$\int_0^\infty e^{tx} . e^{-x/2} x^{(n/2)-1}\, dx$$

$$= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty \exp\left[-\left(\frac{1-2t}{2}\right)x\right] . x^{(n/2)-1}\, dx$$

$$= \frac{1}{2^{n/2}\Gamma(n/2)} \frac{\Gamma(n/2)}{[(1-2t)/2]^{n/2}} \quad \text{[Using Gamma Integral]}$$

$$= (1 - 2t)^{-n/2}, \ |2t| < 1$$

which is the required $m.g.f.$, of a χ^2-variate with $n\,d.f.$

Cumulant Generating Function of χ^2-Distribution

If $X \sim \chi^2(n)$, then

$$K_X(t) = \log M_X(t) = -\frac{n}{2}\log(1-2t)$$

$$= \frac{n}{2}\left[2t + \frac{(2t)^2}{2} + \frac{(2t)^2}{3} + \frac{(2t)^4}{4} + \ldots\ldots\right]$$

$\therefore$ κ_1 = Coefficient of t in $k(t) = n$,

κ_2 = Coefficient of $\frac{t^2}{2!}$ in $K(t) = 2n$,

κ_3 = Coefficient of $\frac{t^3}{3!}$ in $K(t) = 8n$,

and κ_4 = Coefficient of $\frac{t^4}{4!}$ in $K(t) = 48n$

In general, κ_r = Coefficient of $\frac{t^r}{r!}$ in $K(t) = n\,2^{r-1}(r-1)!$

Q. 4. Explain in detail Student's *t*-distribution.

Ans.

Student's *t*-distribution

If $x_1, x_2, \ldots x_n$ is a random sample drawn from a normal population with unknown mean μ and variance σ^2. Also, let

$$\bar{x} = \frac{\Sigma x_i}{n} \text{ and } S^2 = \left[\frac{\Sigma(x_1-\bar{x})^2}{n-1}\right]$$

be the sample mean and pubiased estimate of population variance σ^2, then the statistic

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}}$$

is said to have *t*-distribution with $(n-1)$ degrees of freedom (*d.f.*), *i.e.*, symbolically,

$$t = \frac{\bar{x}-\mu}{S/\sqrt{n}} = \frac{(\bar{x}-\mu)\sqrt{n}}{S} \sim t_{(n-1)d.f.} \quad \ldots(1)$$

Points to be learned

1. The *t*-distribution depends on n, the size of the sample.
2. The sample is drawn from a normal population.
3. As the sample size n increases, *t*-distribution also tends to a standard normal distribution.
4. See that S^2 is different than the sample variance $s^2 = \frac{\Sigma(x_i-\bar{x})^2}{n}$. The relation between S^2 and s^2 is :

$$S^2 = \frac{n}{n-1} s^2 \text{ or } s^2 = \frac{n-1}{n} S^2 \qquad \ldots(2)$$

5. Values of associated probabilities of t-distribution for varying degrees of freedom ($d.f.$) are given in Table-C of the Appendix.

Tests based on *t*-distribution

For small samples ($n < 30$), the following tests of significance are based on t-distribution :

1. Test for an assumed mean, i.e., the test for $H_0 : \mu = \mu_0$.

2. Testing the significance of the difference between two means, i.e., the test $H_0 : \mu_1 = \mu_2$ when (i) the two samples are independent (ii) the two samples are dependent (Paired t-test)

Testing the significance of population or sample mean

In some cases we attempt to test the hypothesis regarding the mean of a population when the sample size is mean ($n < 30$) and population variance σ^2 is unknown. In these cases we draw a random sample of size n from the population and set up the null hypothesis $H_0 : \mu = \mu_0$. The test statistic for testing H_0 is :

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{(n-1)d.f.} \qquad \ldots(3)$$

where, $\bar{x}$ is the sample mean and $S^2 = \frac{\Sigma (x - \bar{x})^2}{n-1}$ is an unbiased estimate of the unknown population variance σ^2. After calculating the value of the statistic in (7), the decision about the acceptance or rejection of H_0 is taken in the following manner :

(a) In case we are testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ (two-sided alternative) at α level of significance, then if :

(i) $|t| \leq t_{(n-1)}(\alpha/2)$; We accept H_0

(ii) $|t| > t_{(n-1)}(\alpha/2)$; We reject H_0

(b) For testing $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ (Right-tailed alternative) at α level of significance, then if

(i) $t \leq t_{(n-1)}(\alpha)$; Accept H_0

(ii) $t > t_{(n-1)}(\alpha)$; Reject H_0

(c) For testing $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$ (Left-tailed alternative) at α level of significance, then if

(i) $t \geq - t_{(n-1)}(\alpha)$; Accept H_0

(ii) $t < - t_{(n-1)}(\alpha)$; Reject H_0

where, $t_{(n-1)}(\alpha)$ is the critical value of t-statistic at α level of significance for $\nu = (n - 1)$ degrees of freedom. These critical values are given in Table-C of the Appendix given at the end of the book.

Important Points

1. Acceptance of $H_0 : \mu = \mu_0$ means that the difference between $\bar{x}$ and μ_0 is insignificant at α level, while rejection of H_0 means that the difference is significant and sample is not a representative of population.

2. If μ is unknown, $\bar{x}$ is an estimator of μ.

3. $(1-\alpha)$ 100% confidence interval for unknown population mean μ is given by

$$\bar{x} = t_{(n-1)}(\alpha/2)\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1)}\}\ (\alpha/2)\frac{S}{\sqrt{n}}$$

Here, $\bar{x} - t_{(n-1)}(\alpha/2)\frac{S}{\sqrt{n}}$ is the lower limit and $\bar{x} + t_{(n-1)}(\alpha/2)\frac{S}{\sqrt{n}}$ is the upper limit of the confidence interval.

In particular :

(i) 90% confidence interval for μ will be :

$$\bar{x} - t_{(n-1)}\,(0.05)\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1)}\,(0.05)\frac{S}{\sqrt{n}}$$

(ii) 95% confidence interval for μ is :

$$\bar{x} - t_{(n-1)}\,(0.025)\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{(n-1)}\,(0.025)\frac{S}{\sqrt{n}}$$

where $t_{(n-1)}\,(0.05)$ and $t_{(n-1)}\,(0.025)$ are critical value of t and can be obtained from Table-C for $(n-1)$ degrees of freedom and for $\alpha = 0.05$ and 0.025 respectively.

4. See that $S^2 = \frac{\Sigma\,(x-\bar{x})^2}{n-1}$ is different than the sample variance $s^2 = \frac{\Sigma\,(x-\bar{x})^2}{n-1}$. The two are related as

$$S^2 = \frac{n}{n-1}s^2 \text{ or } s^2 = \frac{n-1}{nd}S^2$$

If $\bar{x}$ is not an integral value, the value of S^2 should be computed by using the formula

$$S^2 = \frac{1}{n-1}\left[\Sigma\,d^2 - \frac{(\Sigma d)^2}{n}\right],$$ where $d = (x-a)$, a being the assumed mean.

Assumptions of *t*-test

The above t-test is based on the following assumptions :

1. The sample is drawn from a normal population.
2. Sample observations are independent.
3. Sample size is small.
4. Population variance is unknown.

Example. ***Nine items of a sample had the following values :***

45, 47, 50, 52, 48, 47, 49, 53, 51.

Does the mean of nine observations differ significantly from the assumed population mean of 47.5 ?

Solution : **Table Showing Calculations**

S. No.	Sample value x	deviation d = (x–49)	d^2
1	45	–4	16
2	47	–2	4
3	50	+1	1
4	52	3	9
5	48	–1	1
6	47	–2	4
7	49	0	0
8	53	4	16
9	51	2	4
$n = 9$		$\Sigma d = 1$	$\Sigma d^2 = 55$

Here the sample-size $n = 9 < 30$, so it is a small sample case.

For testing the hypothesis $H_0 : \mu = \mu_0 = 47.5$ against $H_1 : \mu \neq \mu_0 = 47.5$, the test statistic is :

$$t = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} \sim t_{(n-1)} \qquad \ldots\text{(i)}$$

where, $\bar{x} = a + \dfrac{\Sigma d}{n} = 49 + \dfrac{1}{9} = 49.11$

$$S^2 = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right] = \frac{1}{9-1}\left[55 - \frac{(1)^2}{9}\right] = \frac{1}{8}\left[55 - \frac{1}{6}\right]$$

$$= \frac{55.89}{8} = 6.86 \therefore S = \sqrt{(6.86)} = 2.62.$$

Putting the value of $n, \bar{x}, \mu_0$ and S in (i), one gets

$$t = \frac{(49.11 - 47.5)\sqrt{9}}{2.62} = \frac{1.61 \times 3}{2.62} = 1.85.$$

The tabulated value of t for $n - 1 = 9 - 1 = 8$ degrees of freedom for 5% level or $t_8(\alpha/2)$, i.e., $t_{8(0.025)}$ is 2.31 (Table C in the appendix). Now, since the calculated value of t is less than the table value, we acept H_0 at 5% level of significance and conclude that sample mean does not differ significantly from population mean.

Testing the significance of the difference between two means

Suppose we have two normal population with means μ_1 and μ_2 respectively. Further suppose that the variances of the two population are unknown but assumed equal, i.e., $\sigma_1{}^2 = \sigma_2{}^2 = \sigma^2$ say where σ^2 is an unknown value. In this situation, we may like to test the null hypothesis

$H_0 : \mu_1 = \mu_2$, *i.e.*, the two population means are equal. For testing H_0, two independent random samples of sizes n_1 and n_2, one from each population, are drawn. Let the observations in the first sample be $x_1, x_2, \ldots, x_{n_1}$ and in the second sample be $y_1, y_2, \ldots, y_{n_2}$.

To test $H_0 : \mu_1 = \mu_2$, we calculate the statistic

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \sim t_{(n_1+n_2-2)} d.f. \qquad \ldots(4)$$

Here, $\bar{x}$ = mean of the first sample.

$\bar{y}$ = mean of the second sample.

$$S^2 = \frac{\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2}{n_1 + n_2 - 2}$$

$$= \frac{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}{n_1 + n_2 - 2},$$

where $\quad S_1^2 = \frac{\Sigma(x - \bar{x})^2}{n_1 - 1}$ and $S_2^2 = \frac{\Sigma(y - \bar{y})^2}{n_2 - 1}$

For computational feasibility, use the following formula for computing S^2 :

$$S^2 = \frac{1}{n_1 + n_2 - 2}\left[\left\{\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1}\right\} + \left\{\Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right\}\right]$$

where, $d_1 = (x - a)$ and $d_2 = (y - b)$, a and b and b being the assumed means.

After computing the value of t statistic in (8), the decision about the acceptance or rejection is taken according to the following rules :

(a) In case we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$ (two-sided), at α level, then if

(i) $|t| \leq t_{(n_1+n_2-2)}(\alpha/2)$; Accept H_0

(ii) $|t| > t_{(n_1+n_2-2)}(\alpha/2)$; Reject H_0.

(b) In case we test $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$ at α level of significance, then if

(i) $t \leq t_{(n_1+n_2-2)}(\alpha)$; Accept H_0

(ii) $t > t_{(n_1+n_2-2)}(\alpha)$; Reject H_0

(c) In case we are testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$ at α level of significance, then if

(i) $t \geq -t_{(n_1+n_2-2)}(\alpha)$ Accept H_0

(ii) $t < -t_{(n_1+n_2-2)}(\alpha)$; Reject H_0.

Here $t_{(n_1+n_2-2)}(\alpha)$ is the critical value of t-statistic at α level of significance for $\nu = (n_1 + n_2 - 2)$ $d.f.$ These critical values are given in Table-C of the Appendix.

Important Remarks

1. Acceptance of $H_0 : \mu_1 = \mu_2$ means that the difference between sample means is insignificant and has arisen due to fluctuations of sampling. In this case, data have no evidence against the hypothesis that the two population means are equal. Rejection of H_0 means the difference between sample means is significant and can be concluded accordingly.

2. $(\bar{x} - \bar{y})$ is an estimator for the difference in population means $(\mu_1 - \mu_2)$.

3. S^2 is an unbiased estimator of common population variance σ^2.

Assumption of the above *t*-test

1. Samples are randomly drawn from normal population.
2. Samples are independent.
3. Sample sizes n_1 and n_2 are small.
4. Population variances are unknown but equal.

Example. *Two kinds of manure were applied to sixteen one acre plots, other conditions remaining the same. The yields are set out below.*

Manure I : 18 20, 36, 50, 49, 36, 34, 49, 41 (9 plots)

Manure II : 29, 28, 26, 35, 30, 44, 49 (7 plots)

Examine the significance of the difference between the mean yields due to the application of different kinds of manure.

Solution : **Table Showing Calculations**

Manure I			Manure II		
Yields x	$d_1=(x-36)$	d_1^2	Yields y	$d_2=(u-34)$	d_2^2
18	−18	324	29	−5	25
20	−10	256	28	−6	36
36	0	0	26	−8	64
50	14	196	35	+1	1
49	13	169	30	−4	16
36	0	0	44	10	100
34	−2	4	49	15	225
49	13	169			
41	5	25			
	$\Sigma d_1=9$	$\Sigma d_1^2=1143$		$\Sigma d_2=3$	$\Sigma d_2^2=467$

Setting up the hypothesis that there is no difference between the mean yields of different kinds of manure, i.e., $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, we calculate the test statistic :

$$t = \frac{\bar{x} - \bar{y}}{s\sqrt{\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \sim t_{(n_1+n_2-2)} \text{ d.f.} \quad \ldots\text{(i)}$$

Here, $\bar{x} = a + \frac{\Sigma d_1}{n_1} = 36 + \frac{9}{9} = 37.0$

$$\bar{y} = b + \frac{\Sigma d_2}{n_2} = 34 + \frac{3}{7} = 34.43$$

$$S^2 = \frac{\Sigma(x - \bar{x})^2 + \Sigma(y - \bar{y})^2}{n_1 + n_2 - 2}$$

$$= \frac{1}{n_1 + n_2 - 2}\left[\left\{\Sigma d_1^2 - \frac{(\Sigma d_1)^2}{n_1}\right\} + \left\{\Sigma d_2^2 - \frac{(\Sigma d_2)^2}{n_2}\right\}\right]$$

$$= \frac{1}{9 + 7 - 2}\left[\left\{1143 - \frac{(9)^2}{9}\right\} - \left\{467 - \frac{(3)^2}{7}\right\}\right]$$

$$= \frac{1134 + 465.71}{14} = 114.27$$

$\therefore S = \sqrt{[114.27]} = 10.69$

Putting the values in (i), the statistic value is

$$t = \frac{37.0 - 34.43}{10.69\sqrt{\left[\frac{1}{9} + \frac{1}{7}\right]}} = \frac{2.57}{10.67 \times 0.504} = \frac{2.57}{5.377} = 0.48.$$

Also; $|t| = 0.48$ and $t_{14}\,(0.025) = 2.14$.

Thus the calculate value is much less than the table values so H_0: is accepted at 5% level of significance and conclude that mean yields due to manure do not differ significantly and the difference is due to fluctuations of sampling.

Testing the significance of the difference between two means (paired t-Test)

In the above testing situation, we assumed that the samples have been randomly drawn from two normal population and they are independent. *However in situations where two samples drawn are not independent, we use* **paired t-test**. Here the observations in the two samples occur in pairs so that the two observations are related in respect of some characteristic. For instance, if we wish to test a new diet using say 15 individuals, then the weights of the individuals recorded before and after completion of the test will form two samples in which observations will be paired. Clearly, the paired observations are recorded on the same individuals or items. Therefore, the two samples will also be of the same size n is view of the paired character of observations and may be put down as

I sample : $x_1 \quad x_2 \quad x_3 \ldots\ldots x_n$

II sample : $y_1 \quad y_2 \quad y_3 \ldots\ldots y_n$.

Using the sample information; we wish to test $H_0 : \mu_1 = \mu_2$ for which the test statistic is

$$t = \frac{\bar{d}}{S_d/\sqrt{n}} \sim t_{(n-1)} \text{ degrees of freedom}$$

Here, $\bar{d} = (\bar{x} - \bar{y}), \bar{x}$ and $\bar{y}$ being the sample means.

$$S_d^2 = \frac{1}{n-1}[\Sigma(d - \bar{d})^2] = \frac{1}{n-1}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right],$$

where $d = (x - y)$.

After computing the value of t-statistic in (9), the decision about acceptance and rejection of H_0 is taken as under :

(a) In case we are testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$, at α level of significance then if

(i) $|t| \leq t_{(n-1)}(\alpha/2)$; Accept H_0

(ii) $|t| > t_{(n-1)}(\alpha/2)$; Reject H_0

(b) Incase of testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 > \mu_2$ at α level of significance, if

(i) $t \leq t_{(n-1)}(\alpha)$; Accept H_0

(ii) $t > t_{(n-1)}(\alpha)$; Reject H_0

(c) In case of testing $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$, if

(i) $t \geq -t_{(n-1)}(\alpha)$; Accept H_0

(ii) $t < -t_{(n-1)}(\alpha)$; Reject H_0

As usual $t_{n-1}(\alpha/2)$ and $t_{(n-1)}(\alpha)$ are critical values of t statistic at α level of significance and for $v = (n-1)$ degrees of freedom, $t_{(n-1)}(\alpha/2)$ is the critical value for a two-tailed test and $t_{(n-1)}(\alpha)$ for a one-tailed test. Significance level for both the test is α.

Assumption of paired *t*-test

1. Two samples are drawn from normal population.
2. Samples are not independent.
3. Population variances are not necessarily equal.
4. Common size of the two samples is small.

Example. *12 school boys were given a test in Mathematics. They were given a month's coaching and a second test, was held at the end of it. The marks recorded are as follows :*

Boys :	1	2	3	4	5	6	7	8	9	10	11	12
Marks in I test :	44	40	61	52	32	44	70	41	67	72	53	72
Marks in II test :	53	38	69	57	46	39	73	48	73	74	60	78

Do not marks given evidence that the students have benefited by the extra coaching ?

Solution. Here we set up the null hypothesis that the boys have no difference in marks before and after the coachings i.e., $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 < \mu_2$ (one-tailed). Here μ_1 and μ_2 stand for the average in the I all II test respectively. In this example, the 12 boys tested on two occasions are the same, as such the scores obtained in the two testing are not independent. So to test the given H_0, we use paired t-test for which the test statistic is : ...(i)

$$t = \frac{\bar{d}\sqrt{n}}{S_d} \sim t_{(n-1)d.f.}$$

Here $d_i = (x_i - y_i), \bar{d} = (\bar{x} - \bar{y})$

and $S_d^2 = \frac{1}{n-1}\Sigma(d_i - \bar{d})^2 = \frac{1}{n-1}\left[\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right]$

Table Showing Calculation

	Marks			
Boys	**I test (x)**	**II test (y)**	**d=(x – y)**	**d²**
1	44	53	–9	81
2	40	38	+2	4
3	61	69	–8	64
4	52	57	–5	25
5	32	46	–14	196
6	44	39	+5	25
7	70	73	–3	9
8	41	48	–7	49
9	67	73	–6	36
10	72	74	–2	4
11	53	60	–7	49
12	72	78	–6	36
			$\Sigma d = -60$	$\Sigma d^2 = 578$

$$\bar{d} = \frac{\Sigma d}{n} = \frac{-60}{12} = -5$$

$$S_d^2 = \frac{1}{n-1}\left\{\Sigma d_i^2 - \frac{(\Sigma d_i)^2}{n}\right\} = \frac{1}{11}\left[578 - \frac{(-60)^2}{12}\right]$$

$$= \sqrt{\left[\frac{278}{11}\right]} = 25.27$$

$\therefore \; S_d = \sqrt{(25.27)} = 5.03$

Putting the values in (i), one gets

$$t = \frac{-5.\sqrt{12}}{5.03} = -3.44 \therefore \; |t| = 3.44.$$

Also the critical value of t for 1 $d.f.$ at 5% level of significance is $t_{11}(0.05) = 1.80$ (for one tailed test). Since calculated value of t is greater than the table value, so H_0 is rejected at 5% level of significance and we conclude that boys marks have increased significantly after coaching or in other words, students have benefitted by the extra coaching. ●

Q. 5. Explain in detail Chi-Square Distribution.

Ans.

Chi-Square Distribution

So for we discussed the sampling distribution of $\bar{x}$, the sample mean. Now let us know about the sampling distribution of sample variance s^2 when samples are randomly drawn from a normal population with mean μ and variance σ^2, Let $x_1, x_2, \ldots\ldots, x_n$ be a sample of size n from the normal population, then the sampling distribution of the statistic

$$\chi^2 = \frac{n s^2}{\sigma^2} = \frac{\Sigma(x - \bar{x})^2}{\sigma^2} \sim \chi^2_{(n-1)} \qquad \ldots(1)$$

Here χ^2 (chi-square) is a statistic which has its distribution as χ^2 with $(n-1)$ degrees of freedom. The distribution was first given by Helmert in 1876 and later was independently derived by **Karl Pearson** in 1900. Another way of defining the χ^2 distribution is as under :

If $X_1, X_2 \ldots\ldots X_n$ are independent normal variates with mean μ and variance σ^2 then

$$Z_i = \frac{X_i - \mu}{\sigma} \; (i = 1, 2, \ldots\ldots n)$$

are independent normal variates with mean 0 and variance 1. Then the sum of squares of the variates $Z_1, Z_2, \ldots Z_n$, i.e.

$$\sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2$$

is distributed as χ^2 with n degrees of freedom. Symbolically,

$$\chi^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right) \sim \chi^2_n$$

Some properties of Chi-square distribution

1. The distribution of χ^2 lies in the first quadrant.

2. The range of χ^2 distributino is from 0 to ∞.

3. It has a unimodal curve.

4. χ^2-distribution has only one parameter n which is its degrees of freedom ($d.f.$).

5. The shape of the curve greatly vary as n varies.

6. The mean and variance of the χ^2 distribution with n $d.f.$ are : mean $= n$; variance $= 2n$.

7. If χ_1^2 and χ_2^2 are two independent χ^2 variates with n_1 and n_2 degrees of freedom respectively, then their sum $\chi_1^2 + \chi_2^2$ will be distributed as chi-square with $n_1 + n_2\,d.f.$ This addition property of chi-square holds good for any number of χ^2 variates. Thus, if $\chi_1^2, \chi_2^2, \ldots\ldots, \chi_k^2$ are k independent χ^2-variates with $n_1, n_2, \ldots\ldots, n_k\ d.f.$ respectively, then their sum $\chi^2 = \sum_{i=1}^{n} \chi_i^2 \sim \chi^2$ with $(n_1 + n_2 + \ldots\ldots + n_k)\ d.f.$

Tests Based on χ^2-Distribution

Following three tests of significance are based on χ^2-distribution.

1. Test of significance for population variance.
2. Testing the goodness of fit.
3. Testing the independence of the two attributes.

These tests have been discussed.

Test of Significance for Population Variance

Let a random sample of size n be drawn from a normal population with mean μ and variance σ_2, μ and σ^2 being unknown. Using the sample information we wish to test the null hypothesis $H_0 : \sigma^2 = \sigma_0^2$, *i.e.*, population variance is σ_0^2. For testing this hypothesis, the test statistic is

$$\chi^2 = \frac{\Sigma(x - \bar{x})^2}{\sigma_0^2} \sim \chi^2_{(n-1)d.f.} \qquad \ldots(2)$$

In terms of s^2 or s^2, the statistic in (2) can be written as

$$\chi^2 = \frac{(n-1)\,S^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$$

After computing the calculated value of χ^2 statistic in (2), the acceptance or rejection of H_0 is done according to the following rules :

(a) In case we test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$ at preassigned value of α, then if

(i) Calculated $\chi^2 \leq \chi^2_{(n-1)}(\alpha)$: Accept H_0 ;

(ii) Calculated $\chi^2 > \chi^2_{(n-1)}(\alpha)$: Reject H_0 :

(b) In case of testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 < \sigma_0^2$ at α level of significance, then if

(i) Calculated $\chi \geq \chi^2_{(n-1)}(1-\alpha)$: Accept H_0 :

(ii) Calculated $\chi^2 < \chi^2_{(n-1)}(1-\alpha)$: Reject H_0 :

Here, $\chi^2_{(n-1)}(\alpha)$ and $\chi^2_{(n-1)}(1-\alpha)$ are critical values of χ^2-statistic and for given α and $v = (n-1)$ degrees of freedom, can be seen from table-D in the appendix. For example, for $n = 10$ and $\alpha = 0.05, \chi^2_{(n-1)}(\alpha) = \chi_9^2\,(0.05) = 16.9$.

Example. ***A random sample of size 25 from a population gives the sample standard derivation to be 9.0. Test the hypothesis that the population standard derivation is 10.5.***

Solution : Given that $n = 25, s^2 = \dfrac{\Sigma\,(x-\bar{x})^2}{n} = (9.0)^2 = 81.$

To test the null hypothesis $H_0 : \sigma^2 = \sigma_0^2 = (10.5)^2$ we calculate the test statistic

$$\chi^2 = \frac{\Sigma\,(x-\bar{x})^2}{n} = (9.0)^2 = 81.$$

To test the null hypothesis $H_0 : \sigma^2 = \sigma_0^2 = (10.5)^2$ we calculate the test statistic

$$\chi^2 = \frac{\Sigma(x-\bar{x})^2}{\sigma_0^2} = \frac{ns^2}{\sigma_0^2}$$

$$= \frac{25 \times 81}{(10.8)^2} = \frac{2025}{116.64} = 17.36.$$

Also the tabulated value of χ^2 for $(n-1) = 24$ degrees of freedom at 5% level is $\chi^2_{24}\,(0.05) = 36.42$.

Now since the calculated value of $\chi^2 = 17.36 < \chi^2_{24}\,(0.05) = 36.42$, H_0 is accepted and we conclude that the population standard deviation may be 10.5.

Testing the Goodness of Fit

So far we have discussed the testing of hypothesis about the population parameters like μ, σ^2 and p. Now we consider a test to determine if a population has a specified theoretical distribution. In other words, here our problem is to test the hypothesis of how closely the observed distribution approximates a particular theoretical distribution. As an example, let us consider a die tossing experiment in which the die is tossed 90 times. Now, under the assumption that the die is fair, the theoretical frequency for each face will be 15. On the other hand, observed frequencies for each face in actual tossing of a die 90 times may differ from theoretical frequencies as shown in the following table :

Observed and theoretical frequencies of 90 tosses of a die

Faces	1	2	3	4	5	6	Total
Theoretical or Expected frequencies	15	15	15	15	15	15	90
Observed Frequencies	18	12	11	19	16	14	90

By comparing the observed frequencies with the corresponding expected frequencies we wish to test whether the differences are likely to occur due to fluctuation of sampling and the die is unbiased or the die is biased and the differences are real.

For explaining the goodness of fit test, let us consider a population which may be partitioned or classified into k classes and let p_i be the probability that an observation belongs to the ith class ($i=1, 2, \ldots, k$) with $\sum_{i=1}^{k} p_i = 1$, Further, let a random sample of size n be drawn from the population. Then suppose o_i is the number of sample observations belonging to ith class and obviously, $\sum_{i=1}^{k} o_i = n$. Also let e_i be the expected frequency in ith class as computed under the null hypothesis. Thus, $e_i = E(o_i) = np_i, i = 1, \ldots, k$. Sum of the expected frequencies too will be n, *i.e.* $\sum_{i=1}^{k} e_i = n$. Then, data can be put as in the following table :

Class	1	2	3	i	......k	Total
Observed freq. o_i	o_1	o_2	o_3	......o_i	......o_k	n
p_i	p_1	p_2	p_3	......p_i	......p_k	n
$e_i = np_i$	e_1	e_2	e_3	......e_i	......n_k	n

Now, a goodness of fit test between observed and expected frequency is based on the statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \qquad \ldots(3)$$

where the sampling distribution of the statistic χ^2 in (1) is χ^2 with $(k-1)$ degrees of freedom.

From χ^2 value in (3), we see that if the observed frequencies (o_i) are close to the expected frequencies (e_i), the χ^2 value will be small, which indicates that the fit is good. One the other hand, if o_i's and e_i's differ considerably, the χ^2 will be large and fit is not good. Thus, acceptance of H_0 : means the fit is good and its rejection means that the fit is not good.

The rejection and acceptance of H_0 at α level of significance depends on the critical value of χ^2-statistics obtained from table-D given in the appendix. This critical value is seen for α level and $\nu = (k-1)$ $d.f.$ and is denoted as $\chi^2_{(k-1)}(\alpha)$. Now if

(i) Calculated $\chi^2 \leq \chi^2_{(k-1)}(\alpha)$; Accept H_0.

(ii) Calculated $\chi^2 > \chi^2_{(k-1)}(\alpha)$; Reject H_0.

Important Remarks

1. In the goodness of fit test observed frequencies in the k class are given while expected frequencies are obtained by assuming the null hypothesis to be true or that the fit is good.

2. Before calculating the value of χ^2 statistic if we observe that the expected frequencies in any class is less than 5, then such frequencies are pooled or combined with adjacent classes. Consequently, the corresponding observed frequencies too are pooled or combined with the adjacent classes and in the process, the number of classes is also reduced. For example, Let us consider the following situation in which we have six classes.

Classes	1	2	3	4	5	6	Total
o_i	3	1	10	15	6	3	38
e_i	1	4	9	18	4	2	38

See that the expected frequencies in first, second, fifth and sixth classes are less than 5, so we combine these frequencies as shown in the adjacent table.

In this way, after combining the frequencies, the number of classes reduces to 3 only.

Pooled Frequencies

Classes	o_i		e_i	
1	3	14	1	14
2	1		4	
3	10		9	
4	15		18	
5	6	9	4	6
6	3		2	
Total	38		38	

Example. *In an experiment on pea breeding, Mendal obtained the following frequencies of seeds :*

315 round and yellow ; 101 wrinkled and yellow ; 108 round and green ; 32 wrinkled and green. Total 556.

Theory predicts that the frequencies would be in the proportion 9 : 3 : 3 : 1. Does the experimental results support the theory ?

Solution. H_0 : Let the Mendalian assumption be true, i.e., the frequencies of peas in the four classes are in the ratio 9 : 3 : 3 : 1.

Thus, the expected frequency of round and yellow seeds

$$E_1 = \frac{9}{16} \times 556 = 313.$$

The expected frequency of wrinkled and yellow seeds

$$E_2 = \frac{3}{16} \times 556 = 104.$$

The expected frequency of round and green seeds

$$E_3 = \frac{3}{16} \times 556 = 104.$$

The expected frequency of wrinkled and green seeds

$$E_4 = \frac{1}{16} \times 556 = 35.$$

Calculation of χ^2

S.No.	Observed frequency (O_i)	Expected frequency E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	315	313	2	4	0.0128
2	101	104	–3	9	0.0865
3	108	104	+4	16	0.1538
4	32	35	–3	9	0.2812
Total	556	556			0.5103

$$\therefore \quad \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = 0.5103$$

Also $\chi_3^2\ (0.05) = 7.815$ [Table-D].

The calculated value of χ_3^2 is much less than the table value.

Thus, we accept H_0 and conclude that there seems a correspondence between theory and experiment.

Testing the independence of two attribute in a contingency table

Suppose we are given a contingency table of order $(r \times s)$ in which N sample observations have been classified. Let O_{ij} be the observed frequency in (i, j)th cell. To test the null hypothesis that the two attributes are independent, we use χ^2-test. First we set up the null hypothesis :

H_0 : Two attributes A and B are independent.

H_1 : Two attributes are dependent or associated.

To test H_0, we calculated the test statistic :

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad \ldots(15)$$

Here, the statistic χ^2 in (4) has χ^2-distribution with $(r-1)(s-1)$ degrees of freedom. The expected frequency corresponding to (i, j)th cell observed frequency, i.e., E_{ij} is obtained by assuming H_0 to be true. Therefore,

$$E_{ij} = E\,(O_{ij}) = \frac{(A_i) \times (B_j)}{N}$$

$$= \frac{\text{Sum of } i\text{th row} \times \text{sum of } j\text{th column}}{\text{Sample size}}$$

$$(i = 1, 2, \ldots, r \text{ and } = 1, 2, \ldots, s)$$

After getting E_{ij}'s or expected frequencies in the cells, χ^2 value is calculated by using the formula in (4). The decision about the acceptance or rejection of H_0 at α level of significance is taken as under :

(i) If calculated $\chi^2 \leq \chi^2_{(r-1)(s-1)}(\alpha)$; Accept H_0

(ii) If calculated $\chi^2 > \chi^2_{(r-1)(s-1)}(\alpha)$; Reject H_0.

$\chi^2_{(r-1)(s-1)}(\alpha)$ is the critical value of χ^2-statistic for $\nu = (r-1)(s-1)$ *d.f.* and for α level of significance. The value of $\chi^2_\nu(\alpha)$ are given in Table-D of the Appendix.

Important Points

1. Acceptance of H_0 means that the two attributes are independent and there is no association between them.

2. Rejection of H_0 leads to the conclusion that H_1 is true and data support the hypothesis that there is some relationship between the two variables.

Following examples will clarify the testing procedure :

Example. ***The following table shows the classification of 4000 workers in a factory, accroding to the disciplinary action taken by the management of the factory and their promotional experience.***

Disciplinary Action	Promotional Experience		Total
	Promoted	Not Promoted	
Non-offenders	300	900	1200
Offenders	100	2700	2800
Total	400	3600	4000

Use χ^2–test to find if there is any evidence to support that there was any association between disciplinary action and the promotional experience of the workers in this factory at 5% level of significance.

Degree of freedom	1	2	3	4	5
χ^2–5%	3.84	5.99	7.82	9.49	11.07

Solution : We form the hypothesis that disciplinary action and promotional experience of the workers are independent. Assuming to be the hypothesis true, we find the expected frequencies for each cell.

Disciplinary Action	Promotional Experience		Total
	Promoted	Not Promoted	
Non-offenders	$\frac{400 \times 1200}{4000} = 120$	$\frac{1200 \times 3600}{4000} = 1080$	1200
Offenders	$\frac{400 \times 2800}{4000} = 280$	$\frac{2800 \times 3600}{4000} = 2520$	2800
Total	400	3600	4000

Let us find the value of χ^2 as

$$= \frac{(300-120)^2}{120} + \frac{(900-1080)^2}{1080} + \frac{(100-280)^2}{280} + \frac{(2700-2520)^2}{2520}$$

$$= \frac{(180)^2}{120} + \frac{(180)^2}{1080} + \frac{(180)^2}{280} + \frac{(180)^2}{2520} - 270 + 30$$

$$+ 115.71 + 12.86 = 428.57.$$

Also table value of χ^2 at 1 d.f. for 5% level of significance is 3.84

Since calculated value of χ^2 is much greater than table value of χ^2, we reject H_0 and conclude that disciplinary action and promotional experience of the workers are dependent.

Q. 6. Discuss F-Distribution in Detail.

Ans.

F-Distribution

Let there be two normal population with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 respectively. The two populations can be symbolically written as $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. Let $x_1, x_2, \ldots, x_{n_1}$ be a sample of size n_1 from the population $N(\mu_1, \sigma_1^2)$ and $y_1, y_2, \ldots, y_{n_2}$ be another sample of size n_2 from the population $N(\mu_2, \sigma_2^2)$. Then according to definition of χ^2, we can form the following two χ^2-statistics :

$$\chi_1^2 = \frac{\Sigma(x-\bar{x})^2}{\sigma_1^2} = \frac{(n_1-1)\,S_1^2}{\sigma_1^2} \text{ and } \chi_2^2 = \frac{\Sigma(y-\bar{y})^2}{\sigma_2^2} = \frac{(n_2-1)\,S_2^2}{\sigma_2^2}$$

where, $S_1^2 = \frac{1}{(n_1-1)}\Sigma(x-\bar{x})^2$ and $S_2^2 = \frac{1}{(n_2-1)}\Sigma(y-\bar{y})^2$ are unbiased estimators of σ_1^2 and σ_2^2 respectively. Now using χ_1^2 and χ_2^2 values, we define the statistic F as :

$$F = \frac{\chi_1^2/(n_1-1)}{\chi_2^2/(n_2-1)} = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \qquad \ldots(5)$$

The sampling distribution of the statistic in (5) is F-distribution with (n_1-1) and (n_2-1) degrees of freedom.

Tests based on F-distribution

The statistical bases of several tests of significance is the F-distribution. Presently, we shall discuss the following two tests based on this distribution :

1. Test of equality of two population variances.
2. Test of equality of several population means.

While test of equality of two population variances is discussed below, **the second testing situation will be discussed in the chapter titled 'Analysis of variance'.**

Test of equality of two population variances

While applying t-distribution for testing $H_0: \mu_1 = \mu_2$, the basic assumption was that the population variances must be equal. F distribution is used for ascertaining the equality of population variances.

Let $x_1, x_2, \ldots, x_{n_1}$ be a sample of size n_1 from the normal population $N(\mu_1, \sigma_1^2)$ and $y_1, y_2, \ldots, y_{n_2}$ be another sample of size n_2 from another normal population $N(\mu_2, \sigma_2^2)$. Using this information we wish to test $H_0: \sigma_1^2 = \sigma_2^2$.

To test H_0 : we calculate the statistic

$$F = \frac{S_1^2}{S_2^2}, (S_1^2 > S_2^2) \qquad \ldots(6)$$

where, $S_1^2 = \frac{1}{n_1 - 1} \Sigma (x - \bar{x})^2$ and $S_2^2 = \frac{1}{n_2 - 1} \Sigma (y - \bar{y})^2$.

The sampling distribution of F-statistic in (6) is F with $v_1 = (n_1 - 1)$ and $v_2 = (n_2 - 1)$ degree of freedom. As a norm, larger variance is taken in the numerator of (6) and the degrees of freedom corresponding to this variance is taken as v_1. The acceptance or rejection of H_0 is decided as under :

(a) In case we test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 > \sigma_2^2$ at α level of significance, then if

(i) Calculated $F \leq F_{v1, v2} (\alpha)$; Accept H_0.

(ii) Calculated $F > F_{v1, v2} (\alpha)$; Reject H_0.

(b) In case we test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 < \sigma_2^2$ at α level, then if

(i) Calculated $F \geq F_{v1, v2} (1 - \alpha)$; Accept H_0.

(ii) Calculated $F < F_{v1, v2} (1 - \alpha)$; Reject H_0.

(c) In case we test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$ at α level of significance, then if

(i) Calculated $F \leq F_{v1, v2} (\alpha/2)$ or $F \geq F_{v1, v2} (1 - \alpha/2)$; Accept H_0.

(ii) Calculated $F > F_{v1, v2} (\alpha/2)$ or $F < F_{v1, v2} (1 - \alpha/2)$; Reject H_0.

where, $F_{v1, v2} (\alpha)$ is the critical value of F-statistic at α level of significance and v_1 and v_2 degrees of freedom. These value can be obtained from table : E and F for $\alpha = 0.05$ and 0.01 respectively.

We consider below some examples.

Example. ***Two samples are drawn from two normal population. From the data information test whether the two samples have the same variance at 5% level of significance.***

Sample 1	60	65	71	74	76	82	85	87		
Sample 2	61	66	67	85	78	63	85	86	88	91

Solution. First we set up the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_2^2 > \sigma_1^2$. Total test H_0, the test statistics is :

$$F = \frac{S_1^2}{S_2^2} \text{ Here, } S_1^2 > S_2^2 \qquad \ldots(i)$$

Table showing calculations

Sample 1 (x)	$(x-\bar{x})$	$(x-\bar{x})^2$	Sample 2 (y)	$(y-\bar{y})$	$(y-\bar{y})^2$
60	−15	225	61	−16	256
65	−10	100	66	−11	121
71	−4	16	67	−10	100
74	−1	1	85	8	64
76	1	1	78	1	1
82	7	49	63	−14	196
85	10	100	85	8	64
87	12	144	86	9	81
			88	11	121
			91	14	196
$\Sigma x = 600$	0	$\Sigma(x-\bar{x})^2 = 636$	$\Sigma y = 770$	0	$\Sigma(y-\bar{y})^2 = 1200$

$$\bar{x} = \frac{\Sigma x}{n_1} = \frac{600}{8} = 75;\ \bar{y} = \frac{\Sigma y}{n_2} = \frac{770}{10} = 77$$

$$S_1^2 = \frac{\Sigma(x-\bar{x})^2}{n_1-1} = \frac{636}{8-1} = 90.86$$

$$S_2^2 = \frac{\Sigma(y-\bar{y})^2}{n_2-1} = \frac{1200}{10-1} = 133.33$$

Since, $S_2^2 > S_1^2$, the F-statistic in (i) will be written as

$$F = \frac{S_2^2}{S_1^2} \sim F_{\nu_1, \nu_2};\ d.f.\ \nu_1 = (n_2 - 1), \nu_2 = (n_1 - 1)$$

$$= \frac{133.33}{90.86} = 1.47.$$

The critical value of F for $\nu_1 = 9$ and $\nu_2 = 7$ and $\alpha = 0.05$ is $F_{9,7}(0.05) = 3.68$ [See Table-E in the Appendix].

Now, since, calculated value of $F < F_{9,7}$ (0.055, H_0 is accepted at 5% level of significance and we conclude that, in the light of given data, the population variances may be assumed equal. ●

Q. 7. Explain the uses of t-distribution in testing the hypothesis.

Or

Discuss the procedure in which the 'student' distribution is generally used to test the significance of the various results obtained from small samples.

Ans. The uses of student distribution in testing the hypothesis are discussed below :

1. Testing the Significance of the Mean of a Random Sample : In determining whether the mean of a sample drawn from a normal population deviates significantly from a stated value (the hypothetical value of the populations mean), when variance of the population is unknown we calculate the statistic :

$$t = \frac{(\bar{X} - \mu)\sqrt{n}}{S}$$

where $\bar{X}$ = the mean of the sample

μ = the actual or hypothetical mean of the population

n = the sample size

S = the standard deviation of the sample

$$S = \sqrt{\frac{\Sigma (X - \bar{X})^2}{n - 1}} \quad \text{or} \quad S = \sqrt{\frac{\Sigma d^2 - n(\bar{d})^2}{n - 1}}$$

$$= \sqrt{\frac{1}{(n-1)}\left[\Sigma d^2 - \frac{(\Sigma d)^2}{n}\right]}$$

where d = deviation from the assumed mean.

If the calculated value of $|t|$ exceeds $t_{0 \cdot 05}$, we say that the difference between $\bar{X}$ and μ is significant at 5% level, if it exceeds $t_{0 \cdot 01}$, the difference is said to be significant at 1% level. If $|t| < t_{0 \cdot 05}$, we conclude that the difference between $\bar{X}$ and μ is not significant and hence the sample might have been drawn from a population with mean $= \mu$.

Fiducial Limits of Population Mean. Assuming that the sample is a random sample from a normal population of unknown mean the 95% fiducial limits of the population mean (μ) are :

$$\bar{X} = \frac{S}{\sqrt{n}} t_{0 \cdot 05}$$

and 99% limits are

$$\bar{X} = \frac{S}{\sqrt{n}} t_{0 \cdot 05}$$

2. Testing Difference Between Means of Two Samples (Independent Samples) : Given two independent random samples of size n_1 and n_2 with means $\bar{X}_1$ and $\bar{X}_2$ and standard deviations S_1 and S_2 we may be interested in testing the hypothesis that the samples come from the same normal population. To carry out the test, we calculate the statistic as follows :

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

where $\bar{X}_1$ = mean of the first sample

$\bar{X}_2$ = mean of the second sample
n_1 = number of observations in the first sample
n_2 = number of observations in the second sample
S = combined standard deviation.

The value of S is calculated by the following formula :

$$S = \sqrt{\frac{\Sigma (X_1 - \bar{X}_1)^2 + \Sigma (X_2 - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

When the actual means are in fraction the deviations should be taken from assumed means. In such a case the combined standard deviation is obtained by applying the following formula :

$$S = \sqrt{\frac{\Sigma (X_1 - A_1)^2 + \Sigma (X_2 - A_2)^2 - n_1 (\bar{X} - A_1)^2 - n_2 (\bar{X}_2 - A_2)^2}{n_1 + n_2 - 2}}$$

A_1 = Assumed mean of the first sample
A_2 = Assumed mean of the second sample
$\bar{X}_1$ = Actual mean of the first sample
$\bar{X}_2$ = Actual mean of the second sample.

The degrees of freedom = $(n_1 + n_2 - 2)$.

When we are given the number of observations and standard deviation of the two samples, the pooled estimate of standard deviation can be obtained as follows :

$$S = \sqrt{\frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}}$$

If the calculated value of t be $> t_{0 \cdot 05}$ $(t_{0 \cdot 01})$, the difference between the sample means is said to be significant at 5% (1%) level of significance otherwise the data are said to be consistent with the hypothesis.

3. Testing Difference between Means of Two Samples (Dependent Samples of Matched Paired Observations) : Two samples are said to be dependent when the elements in one sample are related to those in other in any significant or meaningful manner. In fact, the two samples may consist of pairs of observations made on the same object, individual or, more generally, on the same selected population elements. When samples are dependent they comprise the same number of elementary units. We may carry out some experiment, say, to find out the effect of training on some employees, find out the efficacy of a coaching class or determine whether there is a significant difference in the efficacy of two drugs – one made within the country and another imported. The t-test based on paired observations is defined by the following formula :

$$t = \frac{\bar{d} - 0}{S} \times \sqrt{n} \text{ or } t = \frac{\bar{d}\sqrt{n}}{S}$$

where $\bar{d}$ = the mean of the differences
S = the standard deviation of the differences

The value of S is calculated as follows :

$$S = \sqrt{\frac{\Sigma (d - \bar{d})^2}{n - 1}} \text{ or } \sqrt{\frac{\Sigma d^2 - n (\bar{d})^2}{n - 1}}$$

It should be noted that t is based on $n - 1$ degrees of freedom.

4. Testing the Significance of an Observed Correlation Coefficient : Given a random sample from a bivariate normal population. If we are to test the hypothesis that the correlation coefficient of the population is zero, *i.e.*, the variables in the population are uncorrelated, we have to apply the following test :

$$t = \frac{r}{\sqrt{1 - r^2}} \times \sqrt{n - 2}$$

here t is based on $(n - 2)$ degrees of freedom.

If the calculated value of t exceeds $t_{0\cdot05}$ for $(n - 2)$, *d.f.*, we say that the value of r is significant at 5% level. If $t < t_{0\cdot05}$ the data are consistent with the hypothesis of an uncorrelated population. ●

☛ **Q. 8. The joint p.d.f. $g_{UV}(u, v)$ of the transformed variables U and V is : $g_{UV}(u, v) = f_{XY}(x, y) \mid J \mid$, where $\mid J \mid$ is the modulus value of the Jacobian of transformation and $f(x, y)$ is expressed in terms of u and v.**

Ans. Proof : $P (x < X \leq x + dx, y < Y \leq y + dy)$

$$= P (u < U \leq u + du, v < V \leq + dv)$$

$$\Rightarrow \quad f_{XY}(x, y)\, dx\, dy = g_{UV}(u, v)\, du\, dv$$

$$\Rightarrow \quad g_{UV}(u, v)\, du\, dv = f_{XY}(x, y) \left| \frac{\partial (x, y)}{\partial (u, v)} \right| du\, dv$$

$$\therefore \quad g_{UV}(u, v) = f_{XY}(x, y) \left| \frac{\partial (x, y)}{\partial (u, v)} \right| = f_{XY}(x, y) \mid J \mid$$

●

☛ **Q.9. If X and Y are independent continuous *r.v.'s*, then the *p.d.f.* of $U = X + Y$ is given by :**

$$h(u) = \int_{-\infty}^{\infty} f_X(v)\, f_Y(u - v)\, dv$$

Ans. Let $f_{XY}(x, y)$ be the joint *p.d.f.* of independent continuous *r.v.'s* X and Y and let us make the transformation : $u = x + y, v = x \Rightarrow x = v, y = u - v$.

$$J = \frac{\partial (x, y)}{\partial (u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1$$

Thus the joint *p.d.f.* of *r.v.'s* U and V is given by :

$$g_{UV}(u, v) = f_{XY}(x, y) \mid J \mid = f_X(x) f_Y(y) \mid J \mid$$

(Since X and Y are independent)

$$= f_X(v) f_Y(u - v)$$

The marginal density of U is given by :

$$h(u) = \int_{-\infty}^{\infty} g_{UV}(u, v)\, dv = \int_{-\infty}^{\infty} f_X(v) f_Y(u - v)\, dv$$

Remark : The function $h(.)$ is given a special name and is said to be the convolution of $f_X(.)$ and $f_Y(.)$ and we write $h(.) = f_X(.)^* f_Y(.)$.

Alter : The distribution function $G(.)$ of $U = X + Y$, is given by :

$$G_U(u) = P(U \le u) = \iint_A f(x, y)\, dx\, dy,$$

where $A = \{(x, y) : x + y \le u\}$.

Hence on $A : y \le u - x$.

$$\therefore \quad G_U(u) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{u - x} f(x, y)\, dy \right) dx$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{u} f(x, z - x)\, dz \right) dx$$

$$[x + y = z \Rightarrow dy = dz]$$

$$= \int_{-\infty}^{u} \left(\int_{-\infty}^{\infty} f(x, z - x)\, dx \right) dz$$

(By Fubini's Theorem for non-negative functions)

Differentiating *w.r.* to u, the *p.d.f.* of $U = X + Y$, is given by :

$$g(u) = \int_{-\infty}^{\infty} f(x, u - x)\, dx \qquad ..(3)$$

If X and Y are independent, then

$$g(u) = \int_{-\infty}^{\infty} f_X(x) \,.\, f_y(u - x)\, dx, \qquad ..(4)$$

an expression obtained in (3)

Q. 10. Theorem : Distribution of the Difference of Two Random Variables : If X and Y are independent continuous random variables, then the *p.d.f.* of $U = X - Y$, is given by :

$$h(u) = \int_{-\infty}^{\infty} f(x, x+u)\,dx = \int_{-\infty}^{\infty} f_X(x)\,f_Y(x+u)\,dx \qquad ..(4)$$

Ans. The distribution function $G(.)$ of $U = X - Y$, is given by :

$$G_U(u) = P(U \le u) = \iint\limits_{(x,y)\,:\,x-y \le u} f(x, y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \left(\int_{x-u}^{\infty} f(x, y)\,dy \right) dx, \ [\because x - y \le u \Rightarrow y \ge x - u]$$

Put $\quad x - y = z \Rightarrow x - z = y \Rightarrow dy = -dz;$

when $\quad y = x - u,\ z = x - (x - u) = u .....(*)$

$$\therefore \quad G_U(u) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{u} f(x, x - z)\,dz \right) dx$$

$$= \int_{-\infty}^{u} \left(\int_{-\infty}^{\infty} f(x, x - z)\,dx \right) dz$$

Differentiating *w.r.* to *u*, the *p.d.f.* of $U = X - Y$, is given by :

$$g(u) = \int_{-\infty}^{\infty} f(x, x - u)\,dx \qquad ..(6)$$

$$= \int_{-\infty}^{\infty} f(u + v, v)\,dv \qquad ..(7)$$

If X and y are independent random variables, then the *p.d.f.* of $U = X - Y$, is given by :

$$g(u) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(x - u)\,dx = \int_{-\infty}^{\infty} f_X(u + v)\,f_Y(v)\,dv$$

The range of U is obtained form the given range of X and Y, on using the transformation $(+1)$.

Remark : Similarly proceeding, the *p.d.f.* of $@ = Y - X$ is given by :

$$g_1(w) = \int_{-\infty}^{\infty} f(v, w + v)\,dv \qquad ...(9)$$

$$= \int_{-\infty}^{\infty} f_X(v)\,f_Y(w + v)\,dv, \text{ if } X \text{ and } Y \text{ are independent.} \qquad ..(10)$$

Aliter : Let us transform (X, Y) to (U, V) by the transformation :

$$x - y = u \text{ and } y = v \Rightarrow x = u + v,\ y = v$$

The Jacobian of transformation is given by :

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 0 & 1 \end{vmatrix} = 1.$$

Hence, the joint *p.d.f.* of U and V becomes :

$$g_{UV}(u, v) = f_{XY}(x, y) \mid J \mid = f(u + v, v), 1$$

Integrating *w.r.* to v, the *p.d.f.* of U is given by :

$$g_1(u) = \int_{-\infty}^{\infty} f(u + v, v)\, dv = \int_{-\infty}^{\infty} f(z, z - u)\, dz$$

an expression obtained in (7) or (6).

●

Q. 11. Theorem : Distribution of the Product of two Random Variables : Let X and Y be two independent continuous r.v.'s. Then the p.d.f. of their product U = XY, is given by :

$$h(u) = \int_{-\infty}^{\infty} f_{XY}\left(x, \frac{u}{x}\right) \frac{dx}{\mid x \mid}.$$

Ans. Proof : Let us transform (X, Y) to (U, V) by the transformation : $U = XY$ and $V = X$ so that :

$$u = xy,\ v = x$$

$$\Rightarrow \qquad x = v,\ y = \frac{u}{v} \qquad ..(1)$$

Jacobian of transformation is given by :

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \frac{1}{v} & \frac{-u}{v^2} \end{vmatrix} = \frac{-1}{v}$$

Hence, the joint *p.d.f.* of U and V is given by :

$$g(u, v) = f(x, y) \mid J \mid = f\left(v, \frac{u}{v}\right) \cdot \left|\frac{1}{v}\right|$$

Integrating *w.r.* to v, the *p.d.f.* of $U = XY$, is given by :

$$g(u) = \int_{-\infty}^{\infty} f\left(v, \frac{u}{v}\right) \frac{dv}{\mid v \mid} = \int_{-\infty}^{\infty} f\left(x, \frac{u}{x}\right) \frac{dx}{\mid x \mid} \qquad ..(2)$$

$$= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y\left(\frac{u}{x}\right) \cdot \frac{dx}{\mid x \mid}, \text{ is } X \text{ and } Y \text{ are independent.} \qquad ...(3)$$

The range of U is determined from the given range of X and Y on using the transformation (12).

Q.12. Theorem : Distribution of the Quotient of Two Random Variables : If X and Y are independent continuous random variables, then the *p.d.f.* of the quotient $Z = X/Y$, is given by :

$$g_1(z) = \int_{-\infty}^{\infty} g(vz, v)\,|v|\,dv = \int_{-\infty}^{\infty} f_X(vz) f_Y(v)\,|v|\,dv$$

Ans. Let us transform (X, Y) to (Z, V) by the transformation $Z = X/Y$ and $V = Y$, so that :

$$z = \frac{x}{y},$$

$$v = y$$

$$\Rightarrow \quad x = vz,$$

$$y = v \qquad ..(1)$$

$$J = \begin{vmatrix} \frac{\partial x}{\partial z} & \frac{\partial y}{\partial z} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & 0 \\ z & 1 \end{vmatrix} = v$$

Hence, the joint *p.d.f.* of Z and V is given by :

$$g(z, v) = f(x, y)\,|J| = f(vz, v)\,.\,|v|.$$

Integrating *w.r.* to v, the *p.d.f.* of $Z = X/Y$, is given by :

$$g_1(z) = \int_{-\infty}^{\infty} f(vz, v)\,|v|\,dv \qquad ..(2)$$

$$= \int_{-\infty}^{\infty} f_X(vz) f_Y(v)\,|v|\,dv,$$

if X and Y are independent. ..(3)

The new range of the variable Z is determined from the range of the variables X and Y, on using the transformation (3).

Q. 13. Let $(X, Y)'$ be a two-dimensional non-negative continuous *r.v.*, having the joint density :

$$f(x, y) = \begin{cases} 4xy\,e^{-(x^2 + y^2)}; & x \geq 0, y \geq 0 \\ 0, & \textit{elsewhere} \end{cases}$$

Prove that the density function of $U = \sqrt{X^2 + Y^2}$ is :

$U = \sqrt{X_2 + Y^2}$ is :

$$h(u) = \begin{cases} 2u^3 e^{-u^2}, & 0 \le u < \infty \\ 0, & \textit{elsewhere} \end{cases}$$

. Let us make the transformation : $u = \sqrt{x^2 + y^2}$ and $v = x$

$v \ge 0, u \ge 0$

⇒

and $u \ge v$ or $u \ge v$ or $u \ge 0$ and $0 \le v \le u$.

The Jacobian of transformation J is given by :

$$\frac{1}{J} = \frac{\partial(u, v)}{\partial(x, y)} = \begin{vmatrix} \frac{\partial u}{\partial x} & \frac{\partial v}{\partial x} \\ \frac{\partial u}{\partial y} & \frac{\partial v}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{x}{\sqrt{x^2 + y^2}} & 1 \\ \frac{y}{\sqrt{x^2 + y^2}} & 0 \end{vmatrix} = -\frac{y}{\sqrt{x^2 + y^2}}$$

The joint $p.d.f.$ of U and V is given by :

$$g(u, v) = f(x, y) \mid J \mid = 4xy\, e^{-(x^2 + y^2)} \left| \frac{\sqrt{x^2 + y^2}}{y} \right|$$

$$= 4x \sqrt{x^2 + y^2}\, e^{-(x^2 + 1)}$$

$$= \begin{cases} 4vu \,.\, e^{-u^2} & u \ge 0, 0 \le v \le u \\ 0 & \textit{otherwise} \end{cases}$$

Hence the marginal density function of $U = \sqrt{X^2 + Y^2}$ is :

$$h(u) = \int_0^u g(u, v)\, dv = 4u\, e^{-u^2} \int_0^u v dv = \begin{cases} 2u^3 e^{-u^2}, u \ge 0 \\ 0, \text{elsewhere} \end{cases}$$

Example : Let the $p.d.f.$ of the random variable (x, y) be :

$$f(x, y) = \begin{cases} \alpha^{-2} e^{-(x + y)/\alpha}; & x, y < 0, \alpha < 0 \\ 0, & \textit{elsewhere} \end{cases}$$

Find the distribution of $\frac{1}{2}(X - Y)$.

Solution. Consider the transformation :

$$u = \frac{1}{2}(x - y) \text{ and } v = y$$

$\Rightarrow x = 2u + v$ and $y = v$

The Jacobian of the transformation is :

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = 2$$

Thus, the joint $p.d.f.$ of the random variables (U, V) is given by :

Thus, the joint *p.d.f.* of the random variables (U, V) is given by :

$$g(u, v) = \begin{cases} \frac{2}{\alpha^2} e^{-(2/\alpha)+(u+v)}; & -\infty < u < \infty, \quad v < -2u, \text{ if } u < 0; \\ & v > 0 \text{ if } u \geq 0 \text{ and } \alpha > 0 \\ 0 \quad ; & \textit{elsewhere} \end{cases}$$

The marginal *p.d.f.* of U is given by :

$$gu(u) = \begin{cases} \int_{-2u}^{\infty} \frac{2}{\alpha^2} \exp.\{-(2/\alpha)(u+v)\}\, dv = \frac{1}{\alpha} e^{2u/\alpha}, & u < 0 \\ \int_{-2u}^{\infty} \frac{2}{\alpha^2} \exp.\{-(2/\alpha)(u+v)\}\, dv = \frac{1}{\alpha} e^{-2u/\alpha}, & u \geq 0 \end{cases}$$

Hence $gu(u) = \frac{1}{\alpha} \exp\left\{-\frac{2}{\alpha}|u|\right\} - \infty < u < \infty$

Remark. If $\alpha = 2$, we get

$$g(u) = \frac{1}{2} \exp(-|u|), -\infty < u < \infty,$$

which is *p.d.f.* of standard Laplace Distribution.

Q. 14. Given the joint density function of X and Y as :

$$f(x, y) = \begin{cases} 1/2\, x e^{-y}; 0 < x < 2, y > 0 \\ 0, \qquad \text{elsewhere} \end{cases}$$

find the distribution of $X + Y$

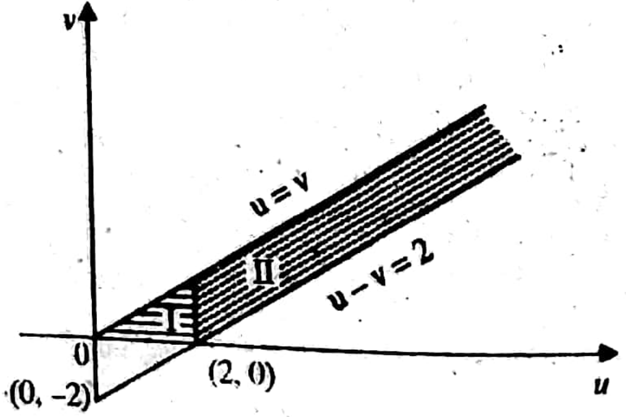
Ans. **Consider the transformation :**

$$u = x + y \text{ and } v = y \Rightarrow y = v, x = u - v$$

The Jacobian of transformation $J = \frac{\partial(x, y)}{\partial(u, v)} = 1$ and the region $0 < x < 2$ and $y > 0$ transforms to $0 < u - v < 2$ and $v > 0$ as shown in the adjoining figure.

The joint density function of U and V is given by :

$$g(u, v) = \frac{1}{2(u-v)} e^{-v}; 0 < v < u, u > 0$$



To find the density of $U = X + Y$, we split the range of U into two parts, (i) $0 < u \leq 2$ (region I) and (ii) $u > 2$ (region II) (which is suggested by the diagram). ●

☛ **Q.15. Theorem. If X_1 and X_2 are two independent χ^2-variates with n_1 and n_2 *d.f.* respectively, then $\dfrac{X_1}{X_2}$ is $\alpha\, \beta_2\left(\dfrac{n_1}{2}, \dfrac{n_2}{2}\right)$ variate.**

Ans. Since X_1 and X_2 are independent χ^2 variates with n_1 and n_2 *d.f.* respectively, their joint probability differential is given by the compound probability theorem as :

$$dP\,(x_1, x_2) = dP_1\,(x_1)\; dP_2\,(x_2)$$

$$= \left[\frac{1}{2^{n_1/2}\,\Gamma\,(n_1/2)} \exp\,(-\,x_1/2)\,(x_1)^{(n_1/2)\,-\,1}\,dx_1\right]$$

$$\times \left[\frac{1}{2^{n_2/n}\,\Gamma\,(n_2/2)} \exp\,(-\,x_2/2)\,(x_2)^{(n_2/2)\,-\,1}\,dx_2\right]$$

$$= \frac{1}{2^{(n_1+n_2)/2}\,\Gamma\,(n_1/2)\,\Gamma\,(n_2/2)} \exp\,\{-\,(x_1+x_2)/2\}$$

$$\times\,(x_1)^{\frac{n_1}{2}-1}\,(x_2)^{\frac{n_2}{2}-1}\,dx_1 dx_2,\; 0 \leq (x_1, x_2) < \infty$$

Let us make the transformation :

$u = x_1/x_2$ and $v = x_2$ so that $x_1 = uv$ and $x_2 = v$.

Jacobian of transformation J is given by :

$$J = \frac{\partial\,(x_1, x_2)}{\partial\,(u, v)} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix} = v$$

Thus the joint distribution of random variables U and V becomes :

$$dG\,(u, v) = \frac{1}{2^{(n_1+n_2)/2}\,\Gamma\,(n_1/2)\,\Gamma\,(n_2/2)}$$

$$\exp\,\{-\,(1+u)\,v/2\} \times (uv)^{\frac{n_1}{2}-}\; v^{\frac{n_2}{2}-1}\,du\,dv,$$

$$= \frac{1}{2^{(n_1+n_2)/2}\,\Gamma\,(n_1/2)\,\Gamma\,(n_2/2)}$$

$$\exp\,\{-\,(1+u)\,v/2\} \times u^{2n_1-1}\,v^{\frac{n_1+n_{22}}{2}-1}\,du\,dv,$$

$$0 \leq (u, v) < \infty$$

Integrating *w.r.* to v over the range 0 to ∞, we get marginal distribution of U as :

$$dG_1(u) = \int_0^\infty dG(u, v)$$

$$= \frac{1}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} u^{(n_1/2)-1} du \times$$

$$\int_0^\infty \exp\left\{-\left(\frac{1+u}{2}\right) v\right\} v^{\{(n_1+n_2)/2\}-1} dv$$

$$= \frac{u^{(n_1/2)-1}}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} \cdot \frac{\Gamma\{(n_1+n_2)/2\}}{[(1+u)/2]^{(n_1+n_2)/2}} du$$

$$= \frac{1}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \cdot \frac{u^{(n_1/2)-1}}{[1+u]^{(n_1+n_2)/2}} du, \ 0 \le u < \infty$$

Hence $U = \frac{X_1}{X_2} \sim \beta_2\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ variate.

Q. 16. If X_1 and X_2 are independent χ^2-variates with n_1 and n_2 d.f. respectively, then $U = \frac{X_1}{X_1 + X_2}$ and $V = X_1 + X_2$ are independently distributed, U as $\beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ variate and V as a χ^2-variate with $(n_1 + n_2)$ d.f.

Ans. Proof. From Theorem 15, we have

$$dP(x_1, x_2) = \frac{1}{2^{(n_1+n_2)/2}\Gamma(n_1/2)\Gamma(n_2/2)} \exp\{-(x_1+x_2)/2\}$$

$$\times (x_1)^{(n_1/2)-1} (x_2)^{(n_2/2)-1} dx_1 dx_2,$$

$$0 \le (x_1, x_2) < \infty$$

Let us transform of u and v defined as follows :

$u = \frac{x_1}{x_1 + x_2}$ and $v = x_1 + x_2$ so, that $x_1 = uv$ and $x_2 = v - x_1 = (1-u)v$

As x_1 and x_2 both range from 0 to ∞, u ranges from 0 to 1 and v from 0 to ∞.

Jacobian of transformation J is : $J = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v$

Q. 17. Write Conditions for the validity of χ^2-test.

Ans. **Conditions for the Validity of χ^2-test :**

χ^2-test is an approximate test for large values of n. For the validity of chi-square test of 'goodness of fit' between theory and experiment, the following conditions must be satisfied :

(i) The sample observations should be independent.

(ii) Constraints on the cell frequencies, if any, should be linear, *e.g.*, $\Sigma n_i = \Sigma \lambda_i$ or $\Sigma O_i = \Sigma E_i$.

(iii) N, the total frequency should be reasonably large, say, greater than 50.

(iv) No theoretical cell frequency should be less than 5. (The chi square distribution is essentially a continuous distribution but it cannot maintain its character of continuity if cell frequency is less than 5.) If any theoretical cell frequency is less than 5, then for the application of χ^2-test, it is pooled with the preceding or succeeding frequency is less than 5, then for the application of χ^2-test, it is pooled with the preceding or succeeding frequency so that the pooled frequency is more than 5 and finally adjust for the *d.f.*, lost in pooling.

It may be noted that the χ^2-test depends only on the set of observed and expected frequencies and on degrees of freedom (*d.f.*). It does not make any assumptions regarding the parent population from which the observations are taken. Since χ^2 defined in does not involve any population parameters, it is termed as a statistic and the test is known as Non-Parmaric Test or Distribution-free Test.

Q. 18. Establish a Relation Between F and χ^2 distribution.

Ans. **Relation Between F and χ^2 Distribution**

In $F(n_1, n_2)$ distribution if we let $n_2 \to \infty$, then $\chi^2 = n_1 F$ follows χ^2-distribution with n_1 *d.f.*

We have

$$f(F) = \frac{(n_1/n_2)^{n_1/2} F^{(n_1/2)-1}}{\Gamma(n_1/2)\,\Gamma(n_2/2)} \cdot \frac{\Gamma[(n_1+n_2)12]}{\left(1+\frac{n_1}{n_2}F\right)^{(n_1+n_2)/2}}, \quad 0 < F < \infty$$

In the limit as $n_2 \to \infty$, we have

$$\frac{\Gamma[(n_1+n_2)/2]}{n_2^{n_1/2}\,\Gamma(n_2/2)} \to \frac{(n_2/2)^{n_1/2}}{n_2^{n_1/2}} = \frac{1}{2^{n_1/2}} \quad ..(1)$$

$$\left[\because \frac{\Gamma(n+k)}{\Gamma(n)} \to n^k \text{ as } n \to \infty.\right]$$

Also $\lim\limits_{n_2 \to \infty} \left(1 + \frac{n_1}{n_2} F\right)^{(n_1+n_2)/2}$

$$= \lim_{n_2 \to \infty} \left[\left(1 + \frac{n_1}{n_2} F\right)^{n_2}\right]^{1/2} \times \lim_{n_2 \to \infty} \left(1 + \frac{n_1}{n_2} F\right)^{n_1/2}$$

$$= \exp(n_1 F/2) = \exp(\chi^2/2) \quad (\because n_1 F = \chi^2) \qquad ..(2)$$

Hence in the limit, on using (1) and (2), the *p.d.f.* of $\chi^2 = n_1 F$ becomes :

$$dP(\chi^2) = \frac{(n_1/2)^{n_1/2} e^{-\chi^2/2}}{\Gamma(n_1/2)} \cdot \left(\frac{\chi^2}{n_1}\right)^{(n_1/2)-1} d\left(\frac{\chi^2}{n_1}\right)$$

$$= \frac{1}{2^{n_1/2}\,\Gamma(n_1/2)} \cdot e^{-\chi^2/2} (\chi^2)^{(n_1/2)-1} d\chi^2,$$

$$0 < \chi^2 < \infty$$

which is the p.d.f. of chi-square distribution with n_1 *d.f.* ●

UNIT 3

Your Syllabus

Calculus of finite differences, operators, separation of symbols, examples and problems. Interpolation formulas with remainder term. Newton's forward and backward formulae. Central difference formulae, Newton's divided difference formulae for interpolation. Lagrange's interpolation formulae.

Compulsory Questions

➲ (1) Drive a relation Between E and Δ and E^{-1} and ∇

Ans. As $Ef(x) = f(x+h)$ and $f(x+h) - f(x) = \Delta f(x)$

$$\therefore \quad \Delta f(x) + f(x) = f(x+h) = Ef(x)$$

$$(\Delta + 1) \equiv E \text{ or } E \equiv 1 + \Delta$$

$$\therefore \quad \boldsymbol{E \equiv 1 + \Delta}$$

$$E^{-1} f(x) = f(x-h)$$

$$E^{-1} f(x) - f(x) = f(x-h) - f(x) = -\nabla f(x)$$

$$\therefore \quad \nabla f(x) = (1 - E^{-1}) f(x)$$

$$\therefore \quad \boldsymbol{E^{-1} \equiv 1 - \nabla}$$

➲ (2) $E\,\Delta \equiv \Delta\,E$ and $E\,\nabla \equiv \nabla\,E \equiv \Delta$

Ans.

$$(E\,\Delta) f(x) = E\,[f(x+h) - f(x)] = f(x+2h) - f(x+h)$$

$$(\Delta\,E) f(x) = \Delta\,[f(x+h)] = f(x+2h) - f(x+h)$$

$$(E\,\nabla) f(x) = E\,[f(x) - f(x-h)] = f(x+h) - f(x) = \Delta f(x)$$

$$(\nabla\,E) f(x) = \nabla\,(f(x+h)) = f(x+h) - f(x) = \Delta f(x)$$

➲ (3) Prove $D \equiv \dfrac{1}{h} \log(1 + \Delta)$

Ans. as $Df(x) = f'(x) = \lim\limits_{h \to 0} \dfrac{f(x+h) - f(x)}{h}$

By Taylor's Theorem

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2!} f''(x) + \ldots$$

$$\therefore \quad \frac{f(x+h) - f(x)}{h} = Df(x) + \frac{h}{2!} D^2 f(x) + \ldots$$

or we can write

$$Ef(x) = f(x) + \frac{h}{\lfloor 1} Df(x) + \frac{h^2}{\lfloor 2} D^2 f(x) + \ldots.$$

$$Ef'(x) = \left(1 + \frac{h}{\lfloor 1} D + \frac{h2}{\lfloor 2} D^2 + \ldots.\right) f(x)$$

$$Ef(x) = e^{hD} f(x)$$

or

$$E = e^{hD} = (1 + \Delta)$$

or

$$D = \frac{1}{h} \log(1 + \Delta)$$

∴

$$D = \frac{1}{h}\left(\Delta - \frac{\Delta^2}{2} + \frac{\Delta^3}{3} + \ldots.\right)$$

or

➲ **(4)** $\delta[f(x)\, g(x)] = \mu f(x)\, \delta g(x) + \mu g(x)\, \delta f(x)$

Ans. L.H.S. $\delta[f(x)\, g(x)]$

$$= f(x + h/2)\, g(x + h/2) - f(x - h/2)\, g(x - h/2)$$

$$\text{R.H.S.} = \frac{1}{2}[f(x + h/2) + f(x - h/2)][g(x + h/2) - g(x - h/2)]$$

$$+ \frac{1}{2}[g(x + h/2) + g(x - h/2)][f(x + h/2) - f(x - h/2)]$$

$$= f(x + h/2)\, g(x + h/2) - f(x - h/2)\, g(x - h/2)$$

$$= \delta[f(x)\, g(x)]$$

$$\therefore \quad \delta[f(x)\, g(x)] = f(x + h)\, \Delta g(x) + g(x)\, \Delta f(x)$$

➲ **(5) Write output the Calculus of Finite Difference and Operator**

Ans. As Δ is forward difference operator, so $\Delta f(a) = f(a + h) - f(a)$ is called first forward difference at point $x = a$, $\Delta f(a + h)$ is called first forward difference at point $x = a + h$

$$\Delta^2 f(a) = \Delta[\Delta f(a)] = \Delta[f(a + h) - f(a)] = \Delta f(a + h) - \Delta f(a)$$

$$\Delta^2 f(a) = f(a + 2h) - 2f(a + h) + f(a)$$

$$\Delta^3 f(a) = \Delta[\Delta^2 f(a)] = \Delta[f(a + 2h) - 2f(a + h) + f(a)]$$

$$= \Delta f(a + 2h) - 2\Delta f(a + h) + \Delta f(a)$$

$$= f(a + 3h) - f(a + 2h) - 2\{f(a + 2h) - f(a + h)\} + \Delta f(a)$$

$$\therefore \Delta^3 f(a) = f(a + 3h) - 3f(a + 2h) + 3f(a + h) - f(a)$$

The most easy method to find the n^{th} difference may be expressed as

$$\Delta^n f(a) = (E - 1)^n f(a) = [E^n - {}^nC_1 E^{n-1} + {}^nC_2 E^{n-2} + \ldots (-1)^n] f(a)$$

$$\Delta^n f(a) = f(a + nh) - {}^nC_1 f(a + \overline{n-1}\, h) + {}^nC_2 f(a + \overline{n-2}\, h) + \ldots (-1)^n f(a)$$

Now $\nabla f(a) = f(a) - f(a - h)$

or $\nabla f(a+h) = f(a+h) - f(a)$

Now $\Delta^2 f(a+2h) = \nabla[\nabla f(a+2h)]$

$$= \nabla[f(a+2h) - f(a+h)]$$

$$= f(a+2h) - f(a+h) - [f(a+h) - f(a)]$$

$$\therefore \nabla^2 f(a+2h) = f(a+2h) - 2f(a+h) + f(a)$$

$$\nabla^n f(x) = \nabla[\nabla^{n-1} f(x)] = \nabla^{n-1} f(x) - \nabla^{n-1} f(x-h)$$

(6) $$u_x - \frac{1}{8}\Delta^2 u_{x-1} + \frac{1\cdot 3}{8\cdot 16}\Delta^4 u_{x-2} - \frac{1\cdot 3\cdot 5}{8\cdot 16\cdot 24}\Delta^3 u_{x-3} + \ldots$$

$$= u_{x+1/2} - \frac{1}{2}\Delta u_{x+1/2} + \frac{1}{4}\Delta^2 u_{x+1/2} - \frac{1}{8}\Delta^3 u_{x+1/2} + \ldots$$

Ans. Taking L.H.S.

$$= u_x - \frac{1}{8}\Delta^2 u_{x-1} + \frac{1\cdot 3}{8\cdot 16}\Delta^4 u_{x-2} - \frac{1\cdot 3\cdot 5}{8\cdot 16\cdot 24}\Delta^2 u_{x-3} + \ldots$$

$$= \left[1 - \frac{1}{8}\Delta^2 E^{-1} + \frac{1\cdot 3}{8\cdot 16}\Delta^4 E^{-2} - \frac{1\cdot 3\cdot 5}{8\cdot 16\cdot 24}\Delta^6 E^{-3} + \ldots\right] u_x$$

$$= \left[1 + \frac{-1}{2}\left(\frac{\Delta^2}{4E}\right) + \frac{\left(-\frac{1}{2}\right)\left(-\frac{1}{2}-1\right)}{2!}\left(\frac{\Delta^2}{4E}\right)^2 + \frac{\left(-\frac{1}{2}\right)\left(-\frac{1}{2}-1\right)\left(-\frac{1}{2}-2\right)}{3!}\left(\frac{\Delta^2}{4E}\right)^3 + \ldots\right] u_x$$

$$= \left[1 + \frac{\Delta^2}{4E}\right]^{-1/2} u_x = \left(\frac{4E}{4E+\Delta^2}\right)^{1/2} u_x = \left[\frac{4E}{4(1+\Delta)+\Delta^2}\right]^{1/2} u_x$$

$$= \left(\frac{4E}{(\Delta+2)^2}\right]^{1/2} u_x = \frac{2E^{1/2}}{2+\Delta} u_x = \left(1 + \frac{\Delta}{2}\right]^{-1/2} u_{x+1/2}$$

$$= \left(1 - \frac{\Delta}{2} + \frac{\Delta^2}{4} + \ldots\right) u_{x+1/2} = u_{x+1/2} - \frac{1}{2}\Delta u_{x+1/2} + \frac{1}{4}\Delta^2 u_{x+1/2} + \ldots$$

(7) Given $u_0 = 3, u_1 = 12, u_2 = 81, u_3 = 2000, u_4 = 100, u_5 = 8$ **find** $\Delta^5 u_0$.

Ans. $\Delta^5 u_0 = (E-1)^5 u_0 = (E^5 - 5E^4 + 10E^3 - 10E^2 + 5E - 1)\, u_0$

$= u_5 - 5u_4 + 10u_3 - 10u_2 + 5u_1 - 40$

$= 8 - 500 + 2000 - 810 + 60 - 3 = 755$

(8) Find $f(6)$ **given that** $f(0) = -3, f(1) = 6, f(2) = 8, f(3) = 12$ **the third difference being constant.**

Ans.

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
0	−3			
		9		
1	6		−7	
		2		9
2	8		2	
		4		
3	12			

$f(6) = E^6 f(0) = (1+\Delta)^6 f(0)$

$= f(0) + 6\Delta f(0) + 15\,\Delta^2 f(0) + 20\,\Delta^3 f(0)$

$= -3 + 6 \times 9 + 15 \times (-7) + 20 \times 9$

$= -3 + 54 - 105 + 180$

$= 126$

(9) Show that $\underset{yz}{\Delta^2} x^3 = x + y + z = f(x, y, z), f(x) = x$

Ans. $f(x,y) = \dfrac{f(y) - f(x)}{y - x} = \dfrac{y^3 - x^3}{y - x} = x^2 + y^2 + xy$

$f(y,z) = y^2 + z^2 + yz$

$f(x,y,z) = \dfrac{f(y,z) - f(x,y)}{z - x} = \dfrac{y^2 + z^2 + yz - x^2 - y^2 - xy}{z - x}$

$= z + x + y$

$= x + y + z$

(10) Find the first and second derivative of the function $y = f(x)$ **tabulated below at the point** $x = 1{\cdot}1$

x :	1	1·2	1·4	1·6	1·8	2·00
$y = f(x)$:	0·0	·1280	·5450	1·2960	2·4320	4·00

Ans. Since the derivative is to be found at $x = 1{\cdot}1$, which is near the begining of given set of values of x in the table and the values of arguments are equally spaced, we shall used Newton's forward formula

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	
1·0	0				
		·1280			
1·2	·1280		·2880		
		·4160		·0480	
1·4	·5440		·3660		0
		·7520		·0480	
1·6	1·2960		·3840		0
		1·1360		·0480	
1·8	2·4320		·4320		
		1·5680			
2·0	4·00				

a = first term = 1
$h = \cdot 2$
Newton's forward formula is

$$f(a+xh) = f(a) + {}^{x}C_1\,\Delta f(a) + {}^{x}C_2\,\Delta^2 f(a) + {}^{x}C_3\,\Delta^3 f(a) + \ldots$$

$$f(a+xh) = f(a) + x\,\Delta f(a) + \frac{x^2 - x}{2}\Delta^2 f(a) + \frac{x^3 - 3x^2 + 2x}{6}\Delta^3 f(a) + \ldots \qquad \ldots(1)$$

Now differentiating it w.r.t. x twice

$$hf'(a+xh) = \Delta f(a) + \frac{(2x-1)}{2}\Delta^2 f(a) + \frac{3x^2 - 6x + 2}{6}\Delta^3 f(a) + \ldots \qquad \ldots(2)$$

$$h^2 f''(a+xh) = \Delta^2 f(a) + \frac{6x - 6}{6}\Delta^3 f(a) + \ldots \qquad \ldots(3)$$

Putting $x = 1/2$, because $a + xh = 1\cdot 1 \Rightarrow x = 1/2$. We get

$$(\cdot 2) f'(1\cdot 1) = \cdot 1280 + \frac{\left(2\cdot\frac{1}{2} - 1\right)}{2}(\cdot 2880) + \frac{3\cdot\frac{1}{4} - 6\cdot\frac{1}{2} + 2}{6}(\cdot 0480)$$

$$(\cdot 2) f'(1\cdot 1) = \cdot 1280 + 0 + \frac{\left(\frac{3}{4} - 3 + 2\right)}{6} \times (\cdot 0480)$$

$$= \cdot 1280 + -\frac{1}{24}(\cdot 04080) = 0\cdot 126$$

$$\therefore\ f'(1\cdot 1) = \frac{0\cdot 126}{0\cdot 2} = 0\cdot 63$$

$$(\cdot 2)^2 f'(1\cdot 1) = (\cdot 2880) + \left(\frac{1}{2} - 1\right) \times (\cdot 0480)$$

$$(\cdot 2)^2 f'(1\cdot 1) = (\cdot 2880) - \frac{1}{2}(\cdot 0480) = 0\cdot 264$$

$$f'(1\cdot 1) = \frac{0\cdot 264}{(\cdot 2)^2} = 6\cdot 6.$$

●

➲ (11) What do you known about Central Difference formula.

Ans. We observe that Newton's divided difference and Lagrange's interpolation formulae are fundamental and are applicable to nearly all cases of interpolation. But in general they do not suit for interpolating values of the function near the middle of a tabulated data. Further more, these formulae do not converge as rapidly as the central difference formulae.

If we want to interpolate the values of $f(x)$ between $x = 0$ and $x = 1$ then it is useful to consider a formula which involves the values of $u_{-3}, u_{-2}, u_{-1}, u_0, u_1, u_2, u_3, \dots$ rather than $u_0, u_1, u_2, u_3, u_4 \dots$. To achieve this we are required to have central difference formulae.

●

➲ (12) Evaluate $\frac{\Delta^2}{E}\sin(x+h) + \frac{\Delta^2 \sin(x+h)}{E\sin(x+h)}$, h being the interval of differencing.

Ans. $\frac{\Delta^2}{E}\sin(x+h) + \frac{\Delta^2 \sin(x+h)}{E\sin(x+h)}$

$$= \frac{(E-1)^2}{E}\sin(x+h) + \frac{(E-1)^2 \sin(x+h)}{\sin(x+2h)}$$

$$= (E - 2 + E^{-1})\sin(x+h) + \frac{(E^2 - 2E + 1)\sin(x+h)}{\sin(x+2h)}$$

$$= [\sin(x+2h) - 2\sin(x+h) + \sin x] + \left[\frac{\sin(x+3h) - 2\sin(x+2h) + \sin(x+h)}{\sin(x+2h)}\right]$$

$$= 2\sin(x+h)[\cos h - 1] + \frac{2\sin(x+2h)(\cos h - 1)}{\sin(x+2h)}$$

[Using trigonometric relations]

$$= 2(\cos h - 1)[\sin(x+h) + 1]$$

●

➲ (13) If $u_x = ax^2 + bx + c$, then show that

$$u_{2n} - {}^nC_1 2u_{2n-1} + {}^nC_2 2^2 u_{n-2} - \dots + (-2)^n u_n = (-1)^n (c - 2an)$$

Ans. Given that

$$u_x = ax^2 + bx + c$$

$$\Rightarrow \quad u_n = an^2 + bn + c.$$

$\therefore$ u_n is a polynomial of degree 2 in n

$$\therefore \quad \Delta^3 u_n = \Delta^4 u_n = \ldots = 0.$$

Let the interval of differencing be equal to 1. Now

$$u_n = an^2 + bn + c$$

$$\Rightarrow \quad \Delta u_n = (a\,(n+1)^2 + b\,(n+1) + c - an^2 - bn - c$$
$$= 2an + a + b$$

and $\Delta^2 u_n = \Delta\,[\Delta u_n]$

$$= 2a\,(n+1) + a + b - 2an - a - b$$
$$= 2a$$

$$\text{LHS} = u_{2n} - {}^nC_1\, 2u_{2n-1} + {}^nC_2\, 2^2\, u_{2n-2} - \ldots$$
$$= [E^n - {}^nC_1\, 2E^{n-1} + {}^nC_2\, 2^2\, E^{n-2} - \ldots]\, u_n$$
$$= (E-2)^n\, u_n$$
$$= (E-1-1)^n\, u_n$$
$$= (\Delta - 1)^n\, u_n \qquad [\therefore \Delta = E - 1]$$
$$= (-1)^n\,(1-\Delta)^n\, u_n$$
$$= (-1)^n\,[1 - {}^nC_1\Delta + {}^nC_2\Delta^2 - \ldots]\, u_n$$
$$= (-1)^n \left[u_n - n\Delta u_n + \frac{n\,(n-1)}{2}\,\Delta^2 u_n - \ldots\right]$$
$$= (-1)^n \left[an^2 + bn + c - n\,(2an + a + b) + \frac{n^2 - n}{2}\, 2a\right]$$
$$= (-1)^n\,(c - 2an)$$
$$= \text{R.H.S.}$$

➲ **(14) Given that $u_x = e^{ax+b}$, find $\Delta^n u_x$.**

Ans. Let h be the interval of differencing

Now $\Delta^n u_x = \Delta^{n-1}\,[\Delta e^{ax+b}]$

$$= \Delta^{n-1}\,[\Delta e^{ax} e^b]$$
$$= \Delta^{n-1} e^b\,[\Delta e^{ax}]$$
$$= e^b\, \Delta^{n-1}\,[e^{ax+h} - e^{ax}]$$
$$= e^b\, \Delta^{n-1}\,[(e^{ah} - 1)\, e^{ax}]$$
$$= e^b\,(e^{ah} - 1)\,\Delta^{n-1} e^{ax}$$
$$= e^b\,(e^{ah} - 1)^2\,\Delta^{n-2} e^{ax}$$
$$\ldots\ldots$$
$$= e^b\,(e^{ah} - 1)^n e^{ax}$$

➲ **(15) Express a given polynomial in factorial notation.**

Ans. Let $f(x)$ be the polynomial of degree n which is to be expressed in factorial notation.

Let $f(x) = A + Bx^{[1]} + Cx^{[2]} + \ldots + Kx^{[n]}$; where $A, B, C, \ldots, K$ are some unknown constants, $K \neq 0$.

Then $\quad \Delta f(x) = B + 2Cx^{[1]} + \ldots\ldots + nKx^{[n-1]},$

$$\Delta^2 f(x) = 2C + 3.2.Dx^{[1]} + \ldots\ldots\ldots + n(n-1)Kx^{[n-2]}$$

$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$$

$$\Delta^n f(x) = Kn.(n-1)\ldots 2.1x^{[0]} = Kn.!$$

Substitution of $x = 0$ gives

$$f(0) = A,$$

$$\Delta f(0) = B,$$

$$\Delta^2 f(0) = C.2! \Rightarrow C = \frac{\Delta^2 f(0)}{2!},$$

$$\Delta^3 f(0) = D.3! \Rightarrow D = \frac{\Delta^3 f(0)}{3!}$$

$$\ldots \ldots \ldots \ldots \ldots$$

$$\ldots \ldots \ldots \ldots \ldots$$

$$\Delta^n f(0) = Kn! \Rightarrow K = \frac{\Delta^n f(0)}{n!}$$

Putting the values of $A, B, C, \ldots K$, we get

$$f(x) = f(0) + \frac{\Delta f(0)}{1!}x^{[1]} + \frac{\Delta^2 f(0)}{2!}x^{[2]} + \frac{\Delta^n f(0)}{n!}x^{[n]}.$$

➲ (16) Give a brief introduction of divided differences.

Ans. The interpolation formulae derived in the preceding chapter are applicable only when the values of the function are given at equidistant intervals of the independent variable or argument.

It is sometimes inconvenient, or even impossible, to obtain values of a function at equidistant values of its argument, and in such cases it is desirable to have interpolation formulae which are applicable when the functional values are given at unequal intervals of the argument. Two such formulae are Newton's formula for unequal intervals of the argument and Lagrange's formula. The former employs diffences, but the latter does not. The differences used in the Newton's formula are called *divided differences*, which are differences obtained in the usual manner and then divided by certain differences of the values of the argument. Hence the name.

●

Long Answer Type Questions

☛ Q. 1. Discuss the Gauss's Forward Formula.

Ans.

Gauss's Forward Formula

$$f(x) = f(x_3) + X_3 \underset{x_1}{\Delta} f(x_3) + X_3 X_4 \underset{x_3, x_4}{\Delta^2} f(x_2) + X_3 X_4 X_2 \underset{x_3, x_4, x_1}{\Delta^3} f(x_2)$$

$$+ X_3 X_4 X_2 X_5 \underset{x_2,x_3x_4x_5}{\overset{4}{\Delta}} f(x_0) + \ldots \qquad \ldots(1)$$

This formula involves $f(x_3)$, $\underset{x_4}{\Delta} f(x_3)$, $\underset{x_3,x_4}{\overset{2}{\Delta}} f(x_2)$, $\underset{x_3x_4x_5}{\overset{3}{\Delta}} f(x_2)$,

$\underset{x_2x_3x_4x_5}{\overset{4}{\Delta}} f(x_1)$, ...

Thus the arguments and their capitals (the usual notation) are brought in the following succession

Arguments $x_3 \;\; x_4 \;\; x_2 \;\; x_5 \;\; x_1 \ldots$

Capitals $1 \;\; X_3 \;\; X_4X_3 \;\; X_4X_3X_2 \;\; X_4X_3X_2X_5 \; \ldots$

If the values of the argument be given at equal intervals of unit length. And if the origin be at point x_3 then

$x_4 = 1, x_5 = 2 \ldots$

$x_2 = -1, x_1 = -2, x_0 = -3$

As from relation of ordinary and divided difference, we have

$$\underset{x_0,x_1}{\overset{n}{\Delta}} f(x_0) = \frac{1}{n!\,h^n} \Delta^n f(x_0) \qquad \text{as } h = 1$$

$$\therefore \underset{x_4}{\Delta} f(x_3) = \Delta f(0), \underset{x_4,x_5}{\overset{2}{\Delta}} f(x_3) = \frac{1}{2!} \Delta^2 f(0)$$

$$\underset{x_3,x_4}{\overset{2}{\Delta}} f(x_2) = \frac{1}{2!} \Delta^2 f(-1), \underset{x_3,x_4,x_5}{\overset{3}{\Delta}} f(x_2) = \frac{1}{3!} \Delta^3 f(-1)$$

Also $X_3 = x - x_3 = x - 0 = x$

$X_4 = x - x_4 = x - 1$

$X_5 = x - 2$

$\therefore \quad X_2 = x - x_2 = x - (-1) = x + 1$

$X_1 = x - x_1 = x - (-2) = x + 2$

$X_0 = x + 3 \ldots\ldots$

The formula (1) reduces to

$$f(x) = f(0) + x\,\Delta f(0) + \frac{x(x-1)}{2!} \Delta^2 f(-1) + \frac{x(x-1)(x-2)}{3!} \Delta^3 f(-1)$$

$$+ \frac{x(x-1)(x+1)(x+2)}{4!} \Delta^4 f(-2) + \ldots \qquad \ldots(2)$$

$$\text{or } f(x) = f(0) + {}^{x}C_1\,\Delta f(0) + {}^{x}C_2\,\Delta^2 f(-1) + {}^{x+1}C_3\,\Delta^3 f(-1)$$

$$+ {}^{x+1}C_4\,\Delta^4 f(-2) + \ldots \qquad \ldots(3)$$

In general (not only for x_3)

This is called Gauss forward formula for equal intervals.

Q. 2. Discuss the Gauss Backward Formula

Gauss Backward Formula

Ans.

When $f(x)$ is expressed in terms of $f(x_3)$ the even differences which lie on the line of $f(x_3)$ and odd differences which lie between the lines of $f(x_2)$ and $f(x_3)$, then the central difference formula so obtained is called Gauss's Backward formula. The backward formula will involve the following terms

$$f(x_3), \underset{x_3}{\Delta} f(x_2), \underset{x_3 x_4}{\Delta} f(x_2), \underset{x_2, x_3, x_4}{\overset{3}{\Delta}} f(x_1), \underset{x_2, x_3, x_4}{\overset{4}{\Delta}} f(x_1) \ldots.$$

From the previous difference table the arguments and capitals are

Argument :	x_3	x_2	x_4	x_1	x_5
Capital :	1	X_3	X_3X_2	$X_3X_2X_4X_3X_2X_4X_1$	

Using Shepard rule, we get

$$f(x) = f(x_3) + X_3 \underset{x_3}{\Delta} f(x_2) + X_3 X_2 \underset{x_3 x_4}{\overset{2}{\Delta}} f(x_2) + X_3 X_2 X_4 \underset{x_2 x_3 x_4}{\overset{3}{\Delta}} f(x_1)$$
$$+ X_3X_2X_4X_1 \underset{x_2 \ldots . x_5}{\overset{4}{\Delta}} f(x_1) + \ldots$$

For equal intervals of unit length $x_3 = 0, x_4 = 1, x_5 = 2, \ldots$ $x_2 = -1, x_1 = -2, x_0 = -3$

Also

$$X_3 = x, X_4 = x - 1, X_5 = x - 2, X_2 = x + 1, X_1 = x + 2 \ldots$$

$$\underset{x_3}{\Delta} f(x_2) = \Delta f(-1); \underset{x_3 x_4}{\overset{2}{\Delta}} f(x_2) = \frac{1}{2!}\Delta^2 f(-1),$$

$$\underset{x_3, x_4, x_2}{\Delta} f(x_1) = \frac{1}{3!}\Delta^3 f(-2) \ldots$$

we have $f(x) = f(0) + x\,\Delta f(-1) + \frac{x(x+1)}{2!}\Delta^2 f(-1)$

$$+ \frac{x(x+1)(x-1)}{3!}\Delta^2 f(-2) + \frac{x(x+1)(x-1)(x+2)}{4!}\Delta^4 f(-2) + \ldots$$

$$f(x) = f(0) + {}^{x}C_1\,\Delta f(-1) + {}^{x+1}C_2\,\Delta^2 f(-1) + {}^{x+1}C_3\,\Delta^3 f(-2)$$
$$+ {}^{x+2}C_4\,\Delta^4 f(-2) + \ldots$$

●

Q.3. Find out the premiums for policies maturing at the age of 46 years from the data given in the table below.

Maturity Age x	45	50	55	60	65
Yearly Premium $f(x)$	2.871	2.404	2.083	1.862	1.712

Ans. The difference table of the given data is as follows :

Maturity Age x	Yearly Premium $f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
45	2.8710				
		− 0.467			
50	2.4040		0.146		
		− 0.321		− 0.046	
55	2.0830		0.100		− 0.017
		− 0.221		− 0.029	
60	1.8630		0.071		
		− 0.150			
65	1.7120				

We want $f(46) = f(a + uh)$, (say), where $a = 45$ and $h = 5$.

$\therefore \quad a + uh = 46 \Rightarrow 45 + u \times 5 = 46$

$\Rightarrow \quad u = \frac{1}{5}$

Substituting $u = \frac{1}{5}$ in Newton's formula for forward interpolation, we get

$$f(46) = f(45) + \frac{1}{5}\Delta f(45) + \frac{1}{5}\left(-\frac{4}{5}\right)\frac{1}{2!}\Delta^2 f(45)$$

$$+ \frac{1}{5}\left(-\frac{4}{5}\right)\left(-\frac{9}{5}\right)\frac{1}{3!}\Delta^3 f(45)$$

$$+ \frac{1}{5}\left(-\frac{4}{5}\right)\left(-\frac{9}{5}\right)\left(-\frac{14}{5}\right)\frac{1}{4!}\Delta^4 f(45)$$

$$= 2.871 + \frac{1}{5}(-0.467) - \frac{2}{25}(0.146) + \frac{6}{125} - (0.046) - \frac{21}{625}(0.017)$$

$$= 2.871 - 0.0934 - .01168 - .002208 = .0005712$$

$$= 2.763 \text{ (approx).}$$

Q. 4. Work out the number of students who obtained marks between 40 and 45 from the table given below

Marks	No. of students
30–40	31
40–50	42
50–60	51
60–70	35
70–80	31

Ans. First we prepare the cumulative frequency table, as given below :

Marks above 30 but less than x	No. of students $f(x)$
40	31
50	73
60	124
70	159
80	190

Now we prepare the difference table for this data.

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
40	31	42	9	−25	37
50	73	51	−16	12	
60	124	35	−4		
70	159	31			
80	190				

We shall find :

$f(45)$ = number of students with marks above 30 but less than 45.

Taking $a + uh = 45$, we get

$$40 + u \times 10 = 45 \Rightarrow u = \frac{1}{2}.$$

Using Newton's formula for forward interpolation, we get

$$f(45) = f(40) + \frac{1}{2}\Delta f(40) + \frac{1}{2}\left(-\frac{1}{2}\right)\frac{\Delta^2 f(40)}{2!}$$

$$+\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{\Delta^3 f(40)}{3!} + \frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\frac{\Delta^4 f(4)}{4!}$$

$$= 31 + \frac{1}{2}\times 42 + \frac{1}{2}\left(-\frac{1}{2}\right)\frac{1}{2!}\times 9 + \frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{1}{3!}(-25)$$

$$+\frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\left(-\frac{5}{2}\right)\frac{1}{4!}\times 37$$

$= 47.868$ (on simplification).

Hence the number of students with marks less than 45 is 47.868 *i.e.*, 48.

But the number of students with marks less than 40 is 31.

Hence the required number of students getting marks marks between 40 and 45 is $= 48 - 31 = 17$.

☛ **Q. 5. From the following**

sin 45° = 0.7071, sin 50° = 0.7660

sin 55° = 0.8192, sin 60° = 0.8660

estimate sin 52° by using any method of interpolation.

Ans. Here, we have

x	45°	50°	55°	60°
$f(x)$	0.7071	0.7660	0.8192	0.8660

The difference table for the given data is as follows :

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^2 f(x)$
45°	0.7071			
		0.0589		
50°	0.7660		− 0.0057	
		0.0532		− 0.0007
55°	0.8192		− 0.0064	
		0.0468		
60°	0.8660			

We want $f(52°) = f(a + uh)$, say.

$$\therefore \quad 52° = a + uh \Rightarrow 52° = 45° + u \times 5° \Rightarrow u = \frac{7}{5} = 1.4.$$

By Newton's forward interpolation formula, we get

$$f(a + uh) = f(a) + u\,\Delta f(a) + \frac{u(u-1)}{2!}\Delta^2 f(a) + \frac{u(u-1)(u-2)}{3!}\Delta^3 f(a)$$

$$\therefore \quad f(52°) = f(45°) + (1.4)\,\Delta f(45°) + \frac{(1.4)(0.4)}{2!}\Delta^2 f(45°) + \frac{(1.4)(0.4)(-0.6)}{3!}\Delta^3 (45°)$$

$$= 0.7071 + 1.4 \times 0.0589 + \frac{(1.4)(0.4)}{2!}(-0.0057) + \frac{(1.4)\,0.4)(-0.6)}{3!} \times (-0.0007)$$

$$= 0.7071 + 0.08246 - 0.001596 + 0.0000392$$

$$= 0.7880032.$$

Thus $\sin 52° = 0.7880032 = 0.7880$ approx.

Q. 6. Estimate the value of $f(7.5)$ from the following data :

x	:	1	2	3	4	5	6	7	8
$f(x)$	:	1	8	27	64	125	216	343	512

Ans. The value to be interpolated lies at the end of the given observations, *i.e.*, near $x = 8$. So in this case Newton's backward formula will be more suitable.

Here $u = \dfrac{x-(a+nh)}{h} = \dfrac{7.5-8}{1} = -0.5.$

To calculate backward differences $\nabla f(a+nh), \nabla^2 f(a+nh), \ldots,$ we prepare the following difference table.

x	$f(x)$	$\nabla f(x)$	$\nabla^2 f(x)$	$\nabla^3 f(x)$
1	1	7	12	6
2	8	19	18	6
3	27	37	24	6
4	64	61	30	6
5	125	91	36	6
6	216	127	42	
7	343	169		
8	512			

Since $\nabla^3 f(x)$ is constant, so we can leave higher order differences.

By Neton's backward interpolation formula,

$$f(a+nh+uh) = f(a+nh) + u\,\nabla f(a+nh) + \frac{u(u+1)}{2!}\nabla^2 f(a+nh)$$
$$+ \frac{u(u+1)(u+2)}{3!}\nabla^3 f(a+nh).$$

$$\therefore \quad f(7.5) = f(8) + (-0.5)\,\nabla f(8) + \frac{(-0.5)(-0.5+1)}{2}\nabla^2 f(8)$$
$$+ \frac{(-0.5)(-0.5+1)(-0.5+2)}{6}\nabla^3 f(8)$$
$$= 512 + (-0.5)\times 169 + \frac{(-0.5)(0.5)}{2}\times 42$$
$$+ \frac{(-0.5)(0.5)(1.5)}{6}\times 6$$
$$= 512 - 84.5 - 5.25 - .375$$
$$= 421.875.$$

Q. 7. Explain the Divided Differences.

Ans. Let $f(x_0), f(x_1), \ldots, f(x_n)$ be the entries corresponding to the arguments $x_0, x_1, \ldots, x_n$ where the intervals $x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1}$ may not be equal, *i.e.*, not necessarily equally spaced. Then the *first divided difference* of $f(x)$ for the arguments x_0, x_1 is defined as

$\dfrac{f(x_1) - f(x_0)}{x_1 - x_0}$ or $\dfrac{f(x_0) - f(x_1)}{x_0 - x_1}$ and is denoted by $f(x_0, x_1)$ or by $\underset{x_1}{\Delta} f(x_0)$.

Similarly the other first divided differences of $f(x)$ for the arguments $x_1, x_2; x_2, x_3; \ldots, x_{n-1}, x_n$ are

$$f(x_1, x_2) = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \underset{x_2}{\Delta} f(x_1)$$

$$f(x_2, x_3) = \frac{f(x_3) - f(x_2)}{x_3 - x_2} = \underset{x_3}{\Delta} f(x_2)$$

$$\ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots$$

$$f(x_{n-1}, x_n) = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} = \underset{x_n}{\Delta} f(x_{n-1})$$

The second divided difference of $f(x)$ for the three arguments x_0, x_1 and x_2 is defined as

$$f(x_0, x_1, x_2) = \frac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0} = \frac{f(x_0, x_1) - f(x_1, x_2)}{x_0 - x_2} = \underset{x_1, x_2}{\Delta^2} f(x_0)$$

The nth divided difference is given by

$$f(x_0, x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n) - f(x_0, x_2, \ldots, x_{n-1})}{x_n - x_0}$$

$$= \frac{f(x_0, x_1, \ldots, x_{n-1}) - f(x_1, x_2, \ldots, x_n)}{x_0 - x_n}$$

$$= \underset{x_1, x_2 \ldots, x_n}{\Delta^n} f(x_0).$$

☛ **Q. 8. The divided differences are symmetrical in all their arguments, that is, the value of any difference is independent of the order of the arguments.**

Ans. We have

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0) - f(x_1)}{x_0 - x_1} = f(x_1, x_0)$$

$$= \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0} = \Sigma \frac{f(x_0)}{x_0 - x_1},$$ showing that $f(x_0, x_1)$ is symmetrical in x_0, x_1.

Again $f(x_0, x_1, x_2) = \dfrac{f(x_1, x_2) - f(x_0, x_1)}{x_2 - x_0}$

$$= \frac{1}{(x_2 - x_0)} \left[\left\{ \frac{f(x_1)}{x_1 - x_2} + \frac{f(x_2)}{x_2 - x_1} \right\} - \left\{ \frac{f(x_0)}{x_0 - x_1} + \frac{f(x_1)}{x_1 - x_0} \right\} \right].$$

$$= \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)}$$

or $f(x_0, x_1, x_2) = \Sigma \dfrac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)}$, showing that $f(x_0, x_1, x_2)$ is symmetrical in x_0, x_1, x_2.

Let us assume similar symmetrical expressions for the $(n-1)$th divided differences *i.e.* let us assume that

$$f(x_0, x_1, \ldots, x_{n-1}) = \frac{f(x_0)}{(x_0 - x_1)\ldots(x_0 - x_{n-1})} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)\ldots(x_1 - x_{n-1})} + \ldots + \frac{f(x_{n-1})}{(x_{n-1} - x_0)(x_{n-1} - x_1)\ldots(x_{n-1} - x_{n-2})}$$

$$= \Sigma \frac{f(x_0)}{(x_0 - x_1)\ldots(x_0 - x_{n-1})},$$

and similar expressions for the other $(n-1)$th divided differences.

Then $f(x_0, x_1, \ldots, x_n) = \dfrac{f(x_0, \ldots, x_{n-1}) - f(x_1, \ldots, x_n)}{x_0 - x_n}$

$$= \frac{1}{(x_0 - x_n)}\left[\left\{\frac{f(x_0)}{(x_0 - x_1)\ldots(x_0 - x_{n-1})} + \frac{f(x_1)}{(x_1 - x_0)\ldots(x_1 - x_{n-1})} + \ldots + \frac{f(x_{n-1})}{(x_{n-1} - x_0)\ldots(x_{n-1} - x_{n-2})}\right\} - \left\{\frac{f(x_1)}{(x_1 - x_2)\ldots(x_1 - x_n)} + \frac{f(x_2)}{(x_2 - x_1)\ldots(x_2 - x_n)} + \ldots + \frac{f(x_n)}{(x_n - x_1)\ldots(x_n - x_{n-1})}\right\}\right]$$

$$= \frac{f(x_0)}{(x_0 - x_1)\ldots(x_0 - x_n)} + \frac{f(x_1)}{(x_1 - x_0)\ldots(x_1 - x_n)} + \ldots \ldots\ldots + \frac{f(x_n)}{(x_n - x_0)\ldots(x_n - x_{n-1})}$$

$$= \Sigma \frac{f(x_0)}{(x_0 - x_1)\ldots(x_0 - x_n)},$$

showing that the nth divided difference $f(x_0, x_1, \ldots, x_n)$ is also symmetrical in $x_0, x_1 \ldots, x_n$ and thus completing the proof of the theorem by mathematical induction.

Thus we see that a divided difference is a symmetrical function of all the arguments involved and it follows that for any function $f(x)$ the value of a divided difference remains unaltered when any two of the arguments involved are interchanged, *i.e.*, the value of a divided difference depends only on the values of the arguments involved and not on the order in which they are taken. Thus

$$f(x_0, x_1) = f(x_1, x_0),$$

$$f(x_0, x_1, x_2) = f(x_1, x_0, x_2) = f(x_2, x_1, x_0) = \ldots, \text{ and so on.}$$

Q. 9. The nth divided difference can be expressed as the quotient of two determinants each of order $n + 1$.

Ans. First consider the function $f(x) = x^n$. The first divided differences of this function are given by

$$f(x_r, x_{r+1}) = \frac{f(x_{r+1}) - f(x_r)}{x_{r+1} - x_r} = \frac{{x_{r+1}}^n - {x_r}^n}{x_{r+1} - x_r}$$

$$= {x_{r+1}}^{n-1} + x_r {x_{r+1}}^{n-2} + \ldots + {x_r}^{n-2} x_{r+1} + {x_r}^{n-1},$$

which is a homogeneous expression of degree $(n - 1)$ in x_r, x_{r+1}.

The second divided differences are given by

$$f(x_r, x_{r+1}, x_{r+2}) = \frac{f(x_r, x_{r+1}) - f(x_{r+1}, x_{r+2})}{x_r - x_{r+2}}$$

$$= \frac{f(x_{r+1}, x_{r+2}) - f(x_r, x_{r+1})}{x_{r+2} - x_r}$$

$$= \frac{1}{(x_{r+2} - x_r)} \Big[\left(x_{r+2}^{n-1} + x_{r+1} x_{r+2}^{n-2} + \ldots + x_{r+1}^{n-2} x_{r+2} + x_{r+1}^{n-1} \right)$$
$$- \left(x_{r+1}^{n-1} + x_r x_{r+1}^{n-2} + \ldots + x_r^{n-2} x_{r+1} + x_r^{n-1} \right) \Big]$$

$$= \frac{x_{r+2}^{n-1} - x_r^{n-1}}{x_{r+2} - x_r} + x_{r+1} \frac{x_{r+2}^{n-2} - x_r^{n-2}}{x_{r+2} - x_r} + \ldots + x_{r+1}^{n-2} \frac{(x_{r+2} - x_r)}{x_{r+2} - x_r}$$

$$= \left(x_{r+2}^{n-2} + \ldots + x_r^{n-2} \right) + x_{r+1} \left(x_{r+2}^{n-3} + \ldots + x_r^{n-3} \right) + \ldots + x_{r+1}^{n-2}$$

which is a homogeneous expression of degree $n - 2$ in x_r, x_{r+1} and x_{r+2}.

By induction it can be shown that $f(x_r, x_{r+1}, \ldots, x_{r+m})$ is a homogeneous expression of degree $n-m$. In particular, the nth divided difference of $f(x) = x^n$ is an expression of degree zero, *i.e.*, it is a constant, and is therefore independent of the values of $x_r, x_{r+1}, x_{r+2}, \ldots, x_{r+n}$.

Since the nth divided differences of x^n are constant, therefore the divided differences of x^n of order greater than n will all be zero.

If $f(x) = ax^n$, where a is a constant, then the nth divided difference of $f(x)$

$= a$ (the nth divided difference of x^n), which is a constant.

Hence if $f(x) = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n$ be a polynomial of degree n, then the nth divided differences of all the terms except $a_0 x^n$ will be zero and so the nth divided difference of the whole polynomial will be constant.

Q. 10. The nth divided difference can be expressed as the quotient of two determinants each of order $n + 1$.

Ans. Let us consider the third divided difference.

We have $f(x_0, x_1, x_2, x_3) = \Sigma \dfrac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)}$

$$= \frac{\Sigma\,[f(x_0)\,.\text{ difference product of } x_1, x_2, x_3]}{\text{difference product of } x_0, x_1, x_2, x_3}.$$

By the theorem of determinants due to Vander-Monde,

$$\text{the difference product of } x_1, x_2, x_3 = \begin{vmatrix} x_1^2 & x_2^2 & x_3^2 \\ x_1 & x_2 & x_3 \\ 1 & 1 & 1 \end{vmatrix}.$$

Therefore,

$$f(x_0, x_1, x_2, x_3) = \Sigma \left[f(x_0) \begin{vmatrix} x_1^2 & x_2^2 & x_3^2 \\ x_1 & x_2 & x_3 \\ 1 & 1 & 1 \end{vmatrix} \right]$$

$$\div \begin{vmatrix} x_0^3 & x_1^3 & x_2^3 & x_3^3 \\ x_0^2 & x_1^2 & x_2^2 & x_3^2 \\ x_0 & x_1 & x_2 & x_3 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

$$\text{or } f(x_0, x_1, x_2, x_3) = \begin{vmatrix} f(x_0) & f(x_1) & f(x_2) & f(x_3) \\ x_0^2 & x_1^2 & x_2^2 & x_3^2 \\ x_0 & x_1 & x_2 & x_3 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

$$\div \begin{vmatrix} x_0^3 & x_1^3 & x_2^3 & x_3^3 \\ x_0^2 & x_1^2 & x_2^2 & x_3^2 \\ x_0 & x_1 & x_2 & x_3 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

Thus the third divided difference has been expressed as the quotient of two determinants each of order 4.

Proceeding in the same way, we can express the higher order differences as the quotient of two determinants each of order one more than the order of the difference.

●

Q. 11. The divided differences can be expressed as the product of multiple integrals i.e.

$$f(x_1, x_2, \ldots, x_n) = \int_0^1 dt_1 \int_0^{t_1} dt_2 \int_0^{t_2} dt_3 \ldots \int_0^{t_{n-2}} f^{n-1}(u_n)\, dt_{n-}$$

where

$u_n = (1 - t_1)x_1 + (t_1 - t_2)x_2 + \ldots + (t_{n-2} - t_{n-1})x_{n-1} + t_{n-1}x_n,$

$t_1, t_2, t_3, \ldots, t_n$ are independent variables and f^{n-1} means the $(n-1)^{th}$ derivative of f.

Ans. First, we shall prove the result for $n = 2$ and $n = 3$.

For $n = 2$, the R.H.S.

$$= \int_0^1 f'(u_2)\, dt_1 = \int_0^1 f'\{(1 - t_1)x_1 + t_1 x_2\}\, dt_1$$

$$= \left[\frac{f\{(1 - t_1)x_1 + t_1 x_2\}}{x_2 - x_1}\right]_0^1$$

$$= \frac{f(x_2) - f(x_1)}{x_2 - x_1} = f(x_1, x_2) = \text{the L.H.S.}$$

Again for $n = 3$, the R.H.S.

$$= \int_0^1 dt_1 \int_0^{t_1} f''(u_3)\, dt_2$$

$$= \int_0^1 dt_1 \int_0^{t_1} f''\{(1 - t_1)x_1 + (t_1 - t_2)x_2 + t_2 x_3\}\, dt_2$$

$$= \int_0^1 dt_1 \left[f'\{(1 - t_1)x_1 + (t_1 - t_2)x_2 + t_2 x_3\}\right]_0^{t_1} \Big/ (x_3 - x_2)$$

$$= \frac{1}{x_3 - x_2}\left[\int_0^1 f'\{(1 - t_1)x_1 + t_1 x_3\}\, dt_1\right.$$

$$- \int_0^1 f'\{(1-t_1)x_1 + t_1x_2\}\,dt_1 \Bigg]$$

$$= \frac{1}{(x_3-x_2)(x_3-x_1)}[f(x_3)-f(x_1)] - \frac{1}{(x_3-x_2)(x_2-x_1)}[f(x_2)-f(x_1)]$$

$$= \frac{f(x_1)}{(x_1-x_2)(x_1-x_3)} + \frac{f(x_2)}{(x_2-x_1)(x_2-x_3)} + \frac{f(x_3)}{(x_3-x_1)(x_3-x_2)}$$

$= f(x_1, x_2, x_3) =$ the L.H.S.

Thus the result is true for $n = 2$ and $n = 3$.

Now let the result be true for n arguments. We shall show that it is also true for $(n+1)$ arguments.

We have $\displaystyle\int_0^{t_{n-1}} f^n(u_{n+1})\,dt_n$

$$= \int_0^{t_{n-1}} f^n\{(1-t_1)x_1 + (t_1-t_2)x_2 + \ldots + (t_{n-1}-t_n)x_n + t_n x_{n+1}\}\,dt_n$$

$$= \frac{1}{x_{n+1}-x_n}\Big[f^{n-1}\{(1-t_1)x_1 + (t_1-t_2)x_2 + \ldots + (t_{n-1}-t_n)x_n + t_n x_{n+1}\}\Big]_0^{t_{n-1}}$$

$$= \frac{1}{x_n - x_{n+1}}\Big[f^{n-1}\{(1-t_1)x_1 + \ldots + (t_{n-2}-t_{n-1})x_{n-1} + t_{n-1}x_n\} - f^{n-1}\{(1-t_1)x_1 + \ldots + (t_{n-2}-t_{n-1})x_{n-1} + t_{n-1}x_{n+1}\}\Big].$$

$$\therefore \int_0^1 dt_1 \int_0^{t_1} dt_2 \ldots \int_0^{t_{n-1}} f^n(u_{n+1})\,dt_n$$

$$= \frac{f(x_1, x_2, \ldots, x_n) - f(x_1, x_2, \ldots, x_{n-1}, x_{n+1})}{x_n - x_{n+1}}$$

$$= \frac{f(x_n, x_2, \ldots, x_1) - f(x_n, x_2, \ldots, x_{n-1}, x_{n+1})}{x_n - x_{n+1}}$$

{interchanging x_1 and x_n}

$$= \frac{f(x_1, x_2, \ldots, x_n) - f(x_2, x_3, \ldots, x_n, x_{n+1})}{x_n - x_{n+1}}$$

$$= f(x_1, x_2, \ldots, x_{n+1}).$$

Thus the result is true for $(n + 1)$ arguments. We have already shown that the result is true for 2 and 3 arguments. Hence by mathematical induction the result is true for n arguments, where n is any +ve integer.

●

☛ Q. 12. Establish Relation Between Divided Differences and Ordinary Differences.

Ans. Let the arguments $x_0, x_1, x_2, \ldots, x_n$ be equally spaced, *i.e.*, $h = x_1 - x_0 = x_2 - x_1 = \ldots = x_n - x_{n-1}$ and let $x = x_0 + uh$. Then

$$f(x_0, x_1) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{1}{h} \Delta f(x_0),$$

$$f(x_0, x_1, x_2) = \frac{f(x_0, x_1) - f(x_1, x_2)}{x_0 - x_2}$$

$$= \frac{\frac{1}{h} \Delta f(x_0) - \frac{1}{h} \Delta f(x_1)}{-2h} = \frac{1}{2!\,h^2} \Delta^2 f(x_0).$$

Similarly

$$f(x_0, x_1, \ldots, x_n) = \frac{1}{n!\,h^n} \Delta^n f(x_0).$$

Substituting these values of the divided differences in Newton's formula

$$f(x_0 + uh) = f(x_0) + \frac{uh}{1!\,h} \Delta f(x_0) + \frac{uh\,(uh - h)}{2!\,h^2} \Delta^2 f(x_0) + \ldots$$

$$+ \frac{uh\,(uh - h)\,(uh - 2h)(uh - \overline{n-1}\,h)}{n!\,h^n} \Delta^n f(x_0)$$

$$= f(x_0) + u\,\Delta f(x_0) + \frac{u\,(u-1)}{2!} \Delta^2 f(x_0) + \ldots$$

$$+ \frac{u\,(u-1)\,(u-2)\ldots(u - \overline{n-1})}{n!} \Delta^n f(x_0)$$

$$= f(x_0) + {}^uC_1\,\Delta f(x_0) + {}^nC_2\,\Delta^2 f(x_0) + \ldots + {}^uC_n\,\Delta^n f(x_0),$$

which is Newton's formula for advancing differences.

●

☛ Q. 13. By means of Newton's divided difference formula, find the values of *f* (8) and *f* (15) from the following table :

x	4	5	7	10	11	13
$f(x)$	48	100	294	900	1210	2028

Ans. The divided difference table is :

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$	$\Delta^4 f(x)$
4	48				
		$\frac{100-48}{5-4}=52$			
5	100		$\frac{97-52}{7-4}=15$		
		$\frac{294-100}{7-5}=97$		$\frac{21-15}{10-4}=1$	
					0
7	294		$\frac{202-97}{10-5}=21$		
		$\frac{900-294}{10-7}=202$		$\frac{27-21}{11-5}=1$	
					0
10	900		$\frac{310-202}{11-7}=27$		
		$\frac{1210-900}{11-10}=310$		$\frac{33-27}{13-7}=1$	
11	1210		$\frac{409-310}{13-10}=33$		
		$\frac{2028-1210}{13-11}=409$			
13	2028				

Using Newton's interpolation formula for unequal intervals, we get

$$f(8)=48+(8-4)\times 52+(8-4)(8-5)\times 15+(8-4)(8-5)(8-7)\times 1$$
$$= 448,$$

and

$$f(15)=48+(15-4)\times 52+(15-4)(15-5)\times 15$$
$$+(15-4)(15-5)(15-7)\times 1$$
$$= 3150.$$

Q. 14. Show that $\underset{bcd}{\Delta^3}\left(\frac{1}{a}\right) = -\frac{1}{abcd}$.

Ans. The divided difference table is as follows :

x	$f(x)$	$\Delta f(x)$	$\Delta^2 f(x)$	$\Delta^3 f(x)$
a	$\frac{1}{a}$			
		$\frac{\frac{1}{b}-\frac{1}{a}}{b-a} = -\frac{1}{ab}$		

b	$\frac{1}{b}$		$(-1)\frac{\frac{1}{bc}-\frac{1}{ab}}{c-a} = (-1)^2\frac{1}{abc}$	
		$\frac{\frac{1}{c}-\frac{1}{b}}{c-b} = -\frac{1}{bc}$		$(-1)^2\frac{\frac{1}{bcd}-\frac{1}{abc}}{d-a} = (-1)^3\frac{1}{abcd}$
c	$\frac{1}{c}$		$(-1)\frac{\frac{1}{dc}-\frac{1}{bc}}{d-b} = (-1)^2\frac{1}{bcd}$	
		$\frac{\frac{1}{d}-\frac{1}{c}}{d-c} = -\frac{1}{dc}$		
d	$\frac{1}{d}$			

Hence, $\underset{bcd}{\Delta^3}\frac{1}{a} = (-1)^3\frac{1}{abcd} = -\frac{1}{abcd}$

Q. 15. Given $\log_{10} 654 = 2\cdot8156$, $\log_{10} 658 = 2\cdot8182$, $\log_{10} 659 = 2\cdot8189$, $\log_{10} 661 = 2\cdot8202$; find $\log_{10} 656$.

Ans. For the values x_0, x_1, x_2, x_3, Lagrange's formula is

$$f(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1) + \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f(x_3).$$

Here $x_0 = 654, x_1 = 658, x_2 = 659, x_3 = 661, x = 656.$

Substituting these values in Lagrange's formula, we get

$$\log_{10} 656 = \frac{(656-658)(656-659-659)(656-661)}{(654-658)(654-659)(654-661)} \times 2\cdot8156$$

$$+\frac{(656-654)(656-659)(656-661)}{(658-654)(658-659)(658-661)}\times 2\cdot8182$$
$$+\frac{(656-654)(656-658)(656-661)}{(659-654)(659-658)(659-661)}\times 2\cdot8189$$
$$+\frac{(656-654)(656-658)(656-659)}{(661-654)(661-658)(661-659)}\times 2\cdot8202.$$

$$\therefore\ \log_{10}656=\frac{(-2)(-3)(-5)}{(-4)(-5)(-7)}\times 2\cdot8156+\frac{(2)(-3)(-5)}{(4)(-1)(-3)}\times 2\cdot8182$$
$$+\frac{(2)(-2)(-5)}{(5)(1)(-2)}\times 2\cdot8189+\frac{(2)(-2)(-3)}{(7)(3)(2)}\times 2\cdot8202$$
$$=0\cdot6033+7\cdot0455-5\cdot6378+0\cdot8058$$
$$=2\cdot8168.$$

Hence the estimated value of $\log_{10}656 = 2\cdot8168$.

Q. 16. The values of y and x are given as below :

x :	5	6	9	11
y :	12	13	14	16

Find the value of y when x = 10.

Ans. Applying Lagrange's formula for $x_0 = 5, x_1 = 6, x_2 = 9, x_3 = 11$ and $x = 10$, we get

$$y_{10}=\frac{(10-6)(10-9)(10-11)}{(5-6)(5-9)(5-11)}\times 12+\frac{(10-5)(10-9)(10-11)}{(6-5)(6-9)(6-11)}\times 13$$
$$+\frac{(10-5)(10-6)(10-11)}{(9-5)(9-6)(9-11)}\times 14+\frac{(10-5)(10-6)(10-9)}{(11-5)(11-6)(11-9)}\times 16$$
$$=\frac{(4)(1)(-1)}{(-1)(-4)(-6)}\times 12+\frac{(5)(1)(-1)}{(1)(-3)(-5)}\times 13$$
$$+\frac{(5)(4)(-1)}{(4)(3)(-2)}\times 14+\frac{(5)(4)(1)}{(6)(5)(2)}\times 16$$
$$=2-4\cdot33+11\cdot66+5\cdot33=14\cdot66.$$

Q. 17. Four equidistant values u_{-1}, u_0, u_1 and u_2 being given, a value is interpolated by Lagrange's formula. Show that it may be written in the form

$$u_x = yu_0 + xu_1 + y\frac{(y^2-1)}{3\,!}\Delta^2u_{-1}+\frac{x(x^2-1)}{3\,!}\Delta^2u_0,$$

where $x + y = 1$.

Ans. We have

$$\Delta^2u_{-1}=(E-1)^2u_{-1}=(E^2-2E+1)\,u_{-1}$$
$$=u_1-2u_0+u_{-1}.$$

Similarly, $\Delta^2 u_0 = (E^2 - 2E + 1)\, u_0$

$= u_2 - 2u_1 + u_0$

Now the R.H.S.

$$= yu_0 + xu_1 + \frac{y(y^2-1)}{3!}\Delta^2 u_{-1} + \frac{x(x^2-1)}{3!}\Delta^2 u_0$$

$$= (1-x)\,u_0 + xu_1 + \frac{(1-x)\,x\,(x-2)}{3!}(u_1 - 2u_0 + u_{-1}) + \frac{x\,(x+1)\,(x-1)}{3!}(u_2 - 2u_1 + u_0)$$

$$= -\frac{x\,(x-1)(x-2)}{6}u_{-1} + \left[-(x-1) + \frac{x\,(x-1)\,(x-2)}{3} + \frac{(x-1)\,x\,(x+1)}{6}\right]u_0 + \left[x - \frac{x\,(x-1)\,(x-2)}{6} - \frac{(x-1)\,x\,(x+1)}{3}\right]u_1 + \frac{(x-1)\,x\,(x+1)}{6}u_2$$

$$= -\frac{x\,(x-1)\,(x-2)}{6}u_{-1} + \frac{(x-2)\,(x-1)\,(x+1)}{2}u_0 - \frac{(x-2)\,x\,(x+1)}{2}u_1 + \frac{(x-1)\,x\,(x+1)}{6}u_2. \quad \ldots(1)$$

For u_{-1}, u_0, u_1, u_2 as known values, by Lagrange's formula, we have

$$u_x = \frac{(x-0)\,(x-1)\,(x-2)}{(-1-0)\,(-1-1)\,(-1-2)}u_{-1} + \frac{(x+1)\,(x-1)\,(x-2)}{(0+1)\,(0-1)\,(0-2)}u_0 + \frac{(x+1)\,(x-0)\,(x-2)}{(1+1)\,(1-0)\,(1-2)}u_1 + \frac{(x+1)\,(x-0)\,(x-1)}{(2+1)\,(2-0)\,(2-1)}u_2$$

$$= -\frac{x\,(x-1)\,(x-2)}{6}u_{-1} + \frac{(x+1)\,(x-1)\,(x-2)}{2}u_0 - \frac{(x+1)\,x\,(x-2)}{2}u_1 + \frac{(x+1)\,x\,(x-1)}{6}u_2. \quad \ldots(2)$$

From (1) and (2), we have

$$u_x = y\,u_0 + x\,u_1 + \frac{y(y^2-1)}{3!}\Delta^2 u_{-1} + \frac{x(x^2-1)}{3!}\Delta^2 u_0.$$

☛ **Q. 18 The values of $f(x)$ are given at a, b and c. Show that the maximum or minimum is attained by**

$$x = \frac{f(a)\,(b^2-c^2) + f(b)\,(c^2-a^2) + f(c)\,(a^2-b^2)}{2\,\{f(a)\,(b-c) + f(b)\,(c-a) + f(c)\,(a-b)\}}.$$

Ans. By Lagrange's formula, for the arguments a, b, c, the function $f(x)$ is given by

$$f(x) = \frac{(x-b)(x-c)}{(a-b)(a-c)} f(a) + \frac{(x-a)(x-c)}{(b-a)(b-c)} f(b) + \frac{(x-a)(x-b)}{(c-a)(c-b)} f(c)$$

$$= \frac{x^2-(b+c)x+bc}{(a-b)(a-c)} f(a) + \frac{x^2-(a+c)x+ac}{(b-a)(b-c)} f(b) + \frac{x^2-(a+b)x+ab}{(c-a)(c-b)} f(c).$$

For maxima or minima, we have $f'(x) = 0$, which gives

$$\frac{2x-(b+c)}{(a-b)(a-c)} f(a) + \frac{2x-(a+c)}{(b-a)(b-c)} f(b) + \frac{2x-(a+b)}{(c-a)(c-b)} f(c) = 0$$

$$\Rightarrow 2x\{(b-c)f(a) + (c-a)f(b) + (a-b)f(c)\} - \{(b^2-c^2)f(a) + (c^2-a^2)f(b) + (a^2-b^2)f(c)\} = 0$$

$$\Rightarrow x = \frac{(b^2-c^2)f(a) + (c^2-a^2)f(b) + (a^2-b^2)f(c)}{2\{(b-c)f(a) + (c-a)f(b) + (a-b)f(c)\}}.$$

Q. 19. If the data are $u_0, u_3, u_4, u_7, u_{11}$ and the interpolation formula is

$$u_x = u_4 + c_1 \underset{3}{\Delta} u_4 + c_2 \underset{4,7}{\Delta^2} u_3 + c_3 \underset{3,4,7}{\Delta^3} u_0 + c_4 \underset{3,4,7,11}{\Delta^4} u_0,$$

find the values of c_1, c_2, c_3 and c_4.

Ans. Newton's divided difference formula is

$$u_x = u(x_0) + (x-x_0)\,u(x_0,x_1) + (x-x_0)(x-x_1)\,u(x_0,x_1,x_2) + (x-x_0)(x-x_1)(x-x_2)\,u(x_0,x_1,x_2,x_3) + (x-x_0)(x-x_1)(x-x_2)(x-x_3)\,u(x_0,x_1,x_2,x_3,x_4).$$

$$\therefore u_x = u_4 + (x-4)\underset{3}{\Delta} u_4 + (x-4)(x-3)\underset{4,7}{\Delta^2} u_3 + (x-7)(x-4)(x-3)\underset{3,4,7}{\Delta^3} u_0 + (x-0)(x-7)(x-4)(x-3)\underset{3,4,7,11}{\Delta^4} u_0$$

Comparing it with the given result, we get

$c_1 = x-4$, $c_2 = (x-4)(x-3)$, $c_3 = (x-7)(x-4)(x-3)$, $c_4 = x(x-7)(x-4)(x-3)$.

Q. 20. Using Lagnange's interpolation formula find a polynomial which passes the points (0, −12), (1, 0), (3, 6), (4, 12).

Ans. We have $x_0 = 0, x_1 = 1, x_2 = 3, x_3 = 4, y_0 = f(x_0) = -12,$

$y_1 = f(x_1) = 0, y_2 = f(x_2) = 6, y_3 = f(x_3) = 12.$

Using Lagrange's interpolation formula we can write

$$f(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} f(x_0) + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} f(x_1) + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)} f(x_3).$$

$$f(x) = \frac{(x-1)(x-3)(x-4)}{(0-1)(0-3)(0-4)} \times (-12) + \frac{(x-0)(x-3)(x-4)}{(1-0)(1-3)(1-4)} \times 0 + \frac{(x-0)(x-1)(x-4)}{(3-0)(3-1)(3-4)} \times (6) + \frac{(x-0)(x-1)(x-3)}{(4-0)(4-1)(4-3)} \times (12)$$

$$= \frac{(x^3-8x^2+19x-12)}{12} \times (12) + \frac{(x^3-5x^2+4x)}{(-6)} \times (6) + \frac{(x^3-4x^2+3x)}{12} \times (12)$$

or $f(x) = x^3 - 7x^2 + 18 - 12$

is the required polynomial.

4

Your Syllabus

Numerical Integration : Derivation of general quadrature formula for equidistant ordinates. Derivation of trapezoidal, Simpson's $1/3^{th}$ and $3\backslash 8^{th}$ rules. Weddle's rule. Real roots of a numerical equation by method of iteration.

Compulsory Questions

➲ **(1) Evaluate $\int_0^{10} \frac{dx}{1+x^2}$ By Weddle's rule.**

Ans. $$\int_1^{10} \frac{1}{1+x^2} dx = \frac{3h}{10} [(y_0 + y_{10}) + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + 2y_6 + y_7$$

$$+ y_8 + 6y_9)$$

$$= \frac{3}{10} [(1 + 9.9009901 \times 10^{-3}) + 5 (0.5) + 0.2 + 6 (0.1)$$
$$+ 0.0588235 + 5 (0.0384615) + 2 (0.027027)$$
$$+ 5 (0.02) + (0.0153846) + 6 (0.0121951)]$$
$$= 1.4410924$$

Now $$\int_0^{10} \frac{1}{1+x^2} dx = \left[\tan^{-1} x\right]_0^{10} = \tan^{-1} 10 = 1.4711277$$

which shows that the value of the integral found by Weddle's rule is the nearest to the actual value than others.

➲ **(2) A curve is drawn to pass through the following points :**

x	1	1.5	2	2.5	3	3.5	4
y	2	2.4	2.7	2.8	3	2.6	2.1

Estimate the area bound by the curve, x-axis and lines $x = 1, x = 4$. Also find the volume of solid generated by revolving this area using Weddle's rule.

Ans. The area bound by the curve, x-axis and lines $x = 1, x = 4$ is given by

$$A = \int_1^4 y\,dx = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + 2y_6]$$

$$= 3\frac{(0.5)}{10}[2 + 5(2.4) + 2.7 + 6(2.8) + 3 + 5(2.6) + 2(2.1)]$$

$$= 8.055 \text{ sq. units}$$

Volume of the solid generated,

$$V = \pi\int_1^4 y^2\,dx = \pi\frac{3h}{10}[y_0^2 + 5y_1^2 + y_2^2 + 6y_3^2 + y_4^2 + 5y_5^2 + 2y_6^2]$$

$$= \pi\, 3\frac{(0.5)}{10}[4 + 5(2.4)^2 + (2.7)^2 + 6(2.8)^2 + 9 + 5(2.6)^2 + 2(2.1)^2]$$

$$= \pi \times 20.8125 = 65.384397 \text{ cubic units.}$$

➲ **(3) A river is 80 ft wide. The depth y ft. at a distance x ft. from one bank is given by the following table calculate the area of cross-section of the river using Simpson's rule.**

x :	0	10	20	30	40	50	60	70	80
y :	0	14	7	9	12	15	14	8	3

Ans. Here $h = 10$

Area of cross-section is $\int_0^8 y\,dx$

$$A = \frac{10}{3}[(0 + 3) + 2(7 + 12 + 14) + 4(4 + 9 + 15 + 8)$$

$$= \frac{10}{3}[3 + 66 + 144]$$

$$= 710 \text{ sq. ft.}$$

➲ **(4) The equation $\sin x = 5x - 2$ can be put as $x = \sin^{-1}(5x - 2)$ and also as $x = \frac{1}{5}(\sin x + 2)$ suggesting two iterative process for its solution which of these if any would succeed and which would fail to give the root in the neighbourhood of 0·5.**

Ans. Let us first consider

$$x = \sin^{-1}(5x - 2)$$

First approximation $x_1 = \sin^{-1}(5x_0 - 2) = \sin^{-1}(5 \times \cdot 5 - 2)$

$$x_1 = \sin^{-1}\frac{1}{2} = \pi/6$$

From $x = \frac{1}{5}\left(\sin\frac{1}{2} + 2\right)$ second relation value of $\sin\frac{1}{2}$ is not known, hence this method fails even for the first approximation

Hence we would succeed from first one

➲ (5) Starting with $x = \cdot 12$, solve $x = \cdot 21 \sin(\cdot 5 + x)$ by using iteration method.

Ans. $x = \cdot 21 \sin(\cdot 5 + x) = \phi(x)$ as $|\phi'(x)| < 1$ hence method is applicable for first approximation

$$x_1 = \cdot 21 \sin(\cdot 5 + x_0)$$

$$x_0 = \cdot 12$$

$$x^{(1)} = x_1 = \cdot 21 \sin(\cdot 5 + \cdot 12) = \cdot 21 \sin(\cdot 62 \text{ rad})$$

$$x_1 = \cdot 21(\cdot 58103) = \cdot 1220$$

$$x^{(2)} = x_2 = \cdot 21 \sin(\cdot 5 + \cdot 1220) = \cdot 21 \sin(\cdot 6220)$$

$$= \cdot 21(0 \cdot 5826) = 0 \cdot 1223$$

$$x^{(3)} = x_3 = \cdot 21 \sin(\cdot 5 + 0 \cdot 1223) = \cdot 21 \sin(0 \cdot 6223)$$

$$= \cdot 21 \times (0 \cdot 5829) = 0 \cdot 1224$$

$$x^{(4)} = x_4 = \cdot 21 \sin(\cdot 5 + 0 \cdot 1224) = \cdot 21 \sin(0 \cdot 6224)$$

$$= \cdot 21(0 \cdot 5830) = 0 \cdot 1224$$

$x^{(3)} = x^{(4)}$, so the required root is $0 \cdot 1224$.

➲ (6) Obtain e^{-1}, where $e = 2 \cdot 71$ starting with $x_0 = 0 \cdot 3$.

Ans. We have $x_{n+1} = x_n(2 - bx_n)$ [formula for inverse of b]

$$b = 2 \cdot 71$$

$$x_1 = x_0(2 - 2 \cdot 71 x_0) = 0 \cdot 3(2 - 2 \cdot 71 \times 0 \cdot 3)$$

$$= 0 \cdot 3(2 - 0 \cdot 819) = 0 \cdot 3543$$

$$x_2 = 0 \cdot 3543(2 - 2 \cdot 71 \times \{0 \cdot 3543\})$$

$$x_2 = 0 \cdot 3543(2 - 0 \cdot 9601) = 0 \cdot 3684$$

$$x_3 = 0 \cdot 3684(2 - 0 \cdot 9983) = 0 \cdot 369003$$

$$x_4 = 0 \cdot 3690(2 - 0 \cdot 9999) = 0 \cdot 369003$$

$x_3 = x_4$. Hence the required root is $0 \cdot 369003$.

➲ (7) Give Table of Comparison Between different Methods.

Ans.

Comparison Between Methods

Method	Formula	Order of convergence	Evaluation of function/ iteration	Reliability of convergence
Bisection	$x_i+1 = \dfrac{x_i + x_i - 1}{2}$ $x_i, x_i - 1$ enclose root	Gain one bit per iteration	1	Guaranteed convergence
False Position	$x_i+1 = \dfrac{x_i - 1 f_i - x_i f_i - 1}{f_i - f_i - 1}$ x_i and $x_i - 1$ enclose root	1	1	Guaranteed convergence
Newton Raphson	$x_i + 1 = x_i - \dfrac{f(x_i)}{f'(x_i)}$	2		Sensitive to starting value convergence fast if starting point near to root.

Secant	$x_i+1 = \dfrac{x_i-1 f_i - x_i f_i-1}{f_i - f_i-1}$ $x_i . x_i - 1$ need not en- close root	1·62	1	No guarantee of convergence
Successive Approximation	$x_i = g\,(x_i - 1)$	1	1	No guarantee of convergence easy to program.

●

➲ (8) Explain the iteration method

Suppose we have an equation

$$f(x) = 0 \qquad ...(1)$$

whose roots are to be determined. The equation (1) can be expressed as

$$x = \phi\,(x). \qquad ...(2)$$

Putting $x = x_0$ in R.H.S. of (2) we get the first approximation

$$x_1 = \phi(x_0).$$

The successive approximations are then given by

$$x_2 = \phi(x_1)$$
$$x_3 = \phi(x_2)$$
$$x_4 = \phi(x_3)$$

.

.

.

$$x_n = \phi(x_{n-1})$$

where the sequence of approximations $x_1, x_2, \ldots, x_n$ always converges to the root of $x = \phi(x)$ if $|\phi'(x)| < 1$, when x is sufficiently close to the exact value of the root. ●

➲ (9) Using tropezoidal rule, calculate the value of the integral $\int_4^{5.2} \log x\, dx$, given

x	4.0	4.2	4.4	4.6	4.8	5.0	5.2
$\log_e x$	1.3863	1.4351	1.4816	1.5260	1.5686	1.6094	1.6486

Compare it with exact value.

Ans. If we take $h = .2$ and $x_0 = 4.0$ and $y = \log_e x$, then we are given $y_0, y_1, y_2, \ldots, y_6$ corresponding to

$$x_0, x_1 = x_0 + h = 4.2, \ldots, x_0 + 6h = 5.2.$$

By Trapezoidal rule, we get

$$\int_4^{5.2} \log_e x\, dx = h\left[(y_0 + y_6)/2 + \sum_{r=1}^{5} y_r\right]$$

$$= (.2)\,[\{(1.3863 + 1.6486)/2\} + \{1.4351 + 1.4816 + 1.5260 + 1.5686 + 1.6094\}]$$

$$= (.2)\,[1.5175 + 7.6207]$$

$$= 1.8276$$

Aslo exact value is given by

$$\int_4^{5.2} \log_e x\, dx \;\left[x \log_e x - x\right]_4^{5.2}$$

$$= [5.2 \log_e 5.2 - 5.2 - 4 \log_e 4 + 4]$$

$$= 5.2 \times 1.6487 - 4 \times 1.3863 - 1.2 = 1.8280$$

Hence error term = 1.8280 − 1.8276 = 0.0004.

We are off in the third decimal place by Trapezoidal rule. ●

➲ **(10) Calculate (upto 3 places of decmila) $\int_2^{10} \frac{dx}{1+x}$ by dividing the range into eight equal parts.**

Ans. Divide the range of integration *i.e.,* (2, 10) into eight equal parts each of width $h = 1$. The values at the end points are given below :

x	: $x_0 = 2$	3	4	5	6	7	8	9	$10 = x_0 + 8h$
$f(x) = \frac{1}{(1+x)}$	: $\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{7}$	$\frac{1}{8}$	$\frac{1}{9}$	$\frac{1}{10}$	$\frac{1}{11}$

By Simpson's '$\frac{1}{3}$' rule, we get

$$\int_{x_0}^{x_0+8h} f(x)\, dx = \frac{h}{3}[f(x_0) + f(x_0 + 8h) + 4\{f(x_0 + h) + \ldots + f(x_0 + 7h)\} + 2\{f(x_0 + 2h) + \ldots + f(x_0 + 6h)\}]$$

$$= \frac{1}{3}\left[\frac{1}{3} + \frac{1}{11} + 4\left\{\frac{1}{4} + \frac{1}{6} + \frac{1}{8} + \frac{1}{14}\right\} + 2\left\{\frac{1}{5} + \frac{1}{7} + \frac{1}{9}\right\}\right]$$

$$= 1{\cdot}299.$$

●

➲ **(11) Evaluate $\int_0^4 e^x\, dx$, by Simpson's rule, using the data $e = 2{\cdot}72$, $e^2 = 7{\cdot}39$, $e^3 = 20{\cdot}09$, $e^4 = 54{\cdot}60$, and compare it with the actual value.**

Ans. Dividing the range (0, 4) into four equal parts at the points $x_0 = 0$, $x_0 + h = 1$, $x_0 + 2h = 2$, $x_0 + 3h = 3$, $x_0 + 4h = 4$, where $h = 1$, we get, by Simpson's '$\frac{1}{3}$' rule,

$$\int_0^4 e^x\, dx = \int_{x_0}^{x_0+4h} f(x)\, dx = \frac{h}{3}[f(x_0) + f(x_0 + 4h) + \ldots$$

$$+ 4\{f(x_0 + h) + (x_0 + 3h)\} + 2f(x_0 + 2h)$$

$$= \frac{1}{3}[1 + 54{\cdot}60 + 4(2{\cdot}72 + 20{\cdot}09) + 2 \times (7{\cdot}39)]$$

$$= \frac{1}{3}[55{\cdot}60 + 91{\cdot}24 + 14{\cdot}78] = 53{\cdot}873$$

The actual value of the integral is

$$\int_0^4 e^x\,dx = e^4 - e_0 = 54{\cdot}60 - 1{\cdot}00 = 53{\cdot}60$$

●

➲ **(12) Calculate $\int_0^{\pi/2} e^{\sin x}\,dx$ correct to four decimal places.**

Ans. Here we shall divide the range $(0, \pi/2 = 90°)$ in three equal parts. So $x_0 = 0, x_1 = \pi/6, x_2 = \pi/3$ and $x_3 = \pi/2$. Obviously the width of each interval is $\pi/6$ *i.e.*, $h = \pi/6$.

The values of the function are given below :

x :	0	$\pi/6$	$\pi/3$	$\pi/2$
$f(x) = e^{\sin x}$:	1	1·64872	2·3632	2·71828

Now by Simpson's '$\frac{3}{8}$' rule, we have

$$\int_0^{\pi/2} e^{\sin x}\,dx = \frac{3h}{8}[f(x_0) + f(x_0 + 3h) + 3\{f(x_0 + h) + f(x_0 + 2h)\}]$$

$$= \frac{3}{8}\cdot\frac{\pi}{6}[1 + 2{\cdot}71828 + 3 \times (1{\cdot}64872 + 2{\cdot}3632)]$$

$$= (\pi/16)[3{\cdot}71828 + 12{\cdot}03576] = {\cdot}091111.$$

●

➲ **(13) Use an approximate integration formula to find the value of $\int_0^6 U_x\,dx$, using the following values of U_x :**

x :	0	1	2	3	4	5	6
U_x :	0·146	0·161	0·176	0·190	0·204	0·217	0·230

Ans. Since the number of equal sub-intervals is six, which is divisible by 3, we shall use Simpson's '$\frac{3}{8}$' rule. We can also use Simpson's '$\frac{1}{3}$' rule.

$$\int_0^6 U_x\,dx = \frac{3h}{8}[U_{x_0} + U_{x_0+6h} + 3\{U_{x_0+h} + U_{x_0+2h} + U_{x_0+4h}$$

$$+ U_{x_0+5h}\} + 2U_{x_0+3h}]$$

$$= (3/8)[0{\cdot}146 + 0{\cdot}230 + 3\{0{\cdot}161 + 0{\cdot}176 + 0{\cdot}204$$

$$+ 0{\cdot}217\} + 2 \times (0{\cdot}190)]$$

$$= (3/8)[0{\cdot}376 + 3 \times {\cdot}758 + 0{\cdot}380] = 1{\cdot}13625.$$

●

➲ **(14) Find the value of $\log_e 2$ from $\int_0^1 \frac{x^2}{1+x^3}\,dx$, using Simpson's '$\frac{1}{3}$' rule, by dividing the range into four equal parts.**

Find the error also.

Ans. We divide the whole range [0, 1] into four equal parts by the ordinates $x = 0, \cdot 25, \cdot 50, \cdot 75, 1\cdot 00$. The values of the function $f(x) = x^2/(1+x^3)$ are given below :

x :	0	·25	·50	·75	1·00
$f(x)$:	0	·06153	·22222	·39560	·500

By Simpson's '$\frac{1}{3}$' rule

$$\int_0^1 \frac{x^2}{1+x^3}\,dx = \int_{x}^{x_0+4h} f(x)\,dx = \frac{h}{3}[f(x_0)+f(x_0+4h)$$

$$+ 4\{f(x_0+h)+f(x_0+3h)\} + 2f(x_0+2h)]$$

$$= \frac{\cdot 25}{3}[0 + \cdot 500 + 4(0\cdot 6153 + \cdot 39560) + 2 \times (\cdot 22222)]$$

$$= \cdot 23108.$$

Now actual value of the integral is

$$\int_0^1 \frac{x^2}{1+x^3}\,dx = \frac{1}{3}\left[\log_e(1+x^3)\right]_0^1 = \frac{1}{3}\log_e 2$$

$$= \frac{1}{3}(\cdot 69315) = \cdot 23105$$

Hence the error $= \cdot 23105 - \cdot 23108 = -\cdot 00003$

➲ **(15) Evaluate $\int_0^1 e^{-x^2}\,dx$ By Simpson's rule.**

Ans. Let $y = f(x) = e^{-x^2}$. Dividing the interval [0, 1] in 10 equal parts, we get $h = \cdot 1$. The tabulated values of y are as follows :

x_i	x_i^2	$f(x_i) = e^{-x_i^2}$
0·0	0·0	1·00000
0·1	0·01	0·99005
0·2	0·04	0·96080
0·3	0·09	0·91393
0·4	0·16	0·85214
0·5	0·25	0·77880
0·6	0·36	0·69768

0·7	0·49	0·61263
0·8	0·64	0·5273
0·9	0·81	0·4449
1·0	1·00	0·36788

Using Simpson's one-third rule, the value of the given integral
$= (\cdot 1/3)\,[(1\cdot 00000 + \cdot 36788) + 4 \times 3\cdot 74031 + 2 \times 3\cdot 03792]$
$= (\cdot 1/3)\,[1\cdot 36788 + 14\cdot 96124 + 6\cdot 07584] = \cdot 746832$ ●

➲ **(16) Find the Moment of Inertia of the area about the x-axis .**

Ans. We know that Moment of Inertia of the area about x-axis is given by

$$\text{M.I.} = (1/3)\,\rho \int_a^b y^3\,dx,$$

where ρ is the mass per unit area.
Here y^3 has the values 1, 64, 512, 64, 1.
Hence by Simpson's rule, (taking $h/4$),

$$\int_a^b y^3\,dx = (1/3)\,(1/4)\,[(1+1) + 4 \times (64+64) + 8 \times 512]$$
$$= 384\cdot 1666.$$ ●

Long Answer Type Questions

☛ **Q. 1. What is Newton's backward interpolation formula ?**

Ans. As for backward interpolation we write

$$P_n = A_0 + A_1\,(x - \overline{a+nh}) + A_2\,(x - \overline{a+nh})\,(x - \overline{a+n-1\,h})$$
$$+ A_3\,(x - \overline{a+nh})\,(x - \overline{a+n-1\,h})\,(x - \overline{a+n-2h}) + \ldots$$
$$+ \ldots + A_n\,(x - \overline{a+nh})\,(\quad)\,\ldots\ldots\,(x - \overline{a+h})$$

where $A_0, A_1, \ldots, A_n$ are constant so chosen such that

$$P_n\,(a+nh) = f\,(a+nh), \ldots, P_n\,(a) = f\,(a)$$

putting in (1), $x = a + nh, a + \overline{n-1}\,h$, we get

$$A_0 = f\,(a+nh),\ f\,(a + \overline{n-1}\,h) = A_0 + A_1\,(-h)$$

$$\text{or } A_1 = \frac{f\,(a+nh) - f\,(a + \overline{n-1}\,h)}{h} = \frac{\nabla f\,(a+nh)}{h}$$

Similarly $A_2 = \dfrac{1}{2!\,h^2}\nabla^2 f(a+nh) \ldots A_n = \dfrac{1}{n!\,h^n}\nabla^n (a+nh)$

$$\therefore P_n(x) = f(a+nh) + \frac{\nabla f(a+nh)}{h} + \frac{1}{2!\,h^2}\nabla^2 f(a+nh) + \ldots + \frac{1}{n!\,h^n}\nabla^n f(a+nh)$$

Putting $u = \dfrac{x-(a+nh)}{h}$

or $\qquad x = (a+nh+hu)$

$$P_n(x) = P_n(a+nh+hu) = f(a+nh) + u\,\nabla f(a+nh) + \frac{u(u+1)}{2!}\nabla^2 f(a+nh) + \ldots\ldots + \frac{u(u+1)\ldots\ldots(u+n-1)}{n!}\nabla^n f(a+nh)$$

This formula is used when the required value is near the last value.

☛ **Q. 2. Derive general quadrature formula for numerical integration.**

Ans. Let $I = \int_a^b y\,dx$, where $y = f(x)$ takes the values $y_0, y_1, y_2, \ldots, y_n$ for $x_0, x_1, x_2, \ldots, x_n$. We divide the interval (a, b) into n equal parts of width h, so that

$a = x_0, x_1 = x_0 + h, x_2 = x_0 + 2h, \ldots, x_n = x_0 + nh = b$

Then $I = \displaystyle\int_{x_0}^{x_0+nh} f(x)\,dx$

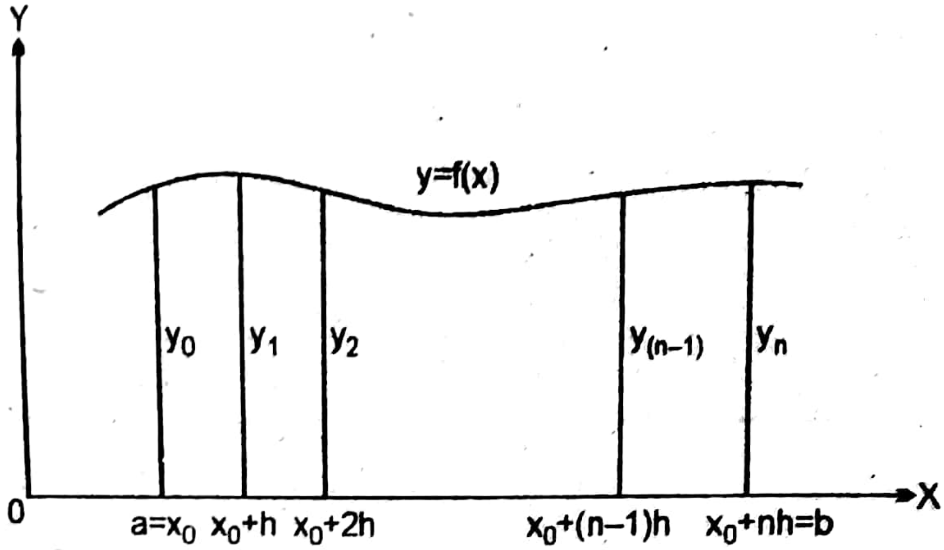
Putting $x = x_0 + ph$, so that $dx = h\,dp$, in above, we get

$$I = h\int_0^n f(x_0+ph)\,dp = h\int_0^n y_p\,dp$$

Now replacing y_p by Newton's forward interpolation formula, we get

$$I = h\int_0^n \left[y_0 + p\,\Delta y_0 + \frac{p(p-1)}{2!}\Delta^2 y_0 + \frac{p(p-1)(p-2)}{3!}\Delta^3 y_0 + \frac{p(p-1)(p-2)(p-3)}{4!}\Delta^4 y_0 + \frac{p(p-1)(p-2)(p-3)(p-4)}{5!}\Delta^5 y_0 \right.$$
$$\left. \frac{p(p-1)(p-2)(p-3)(p-4)(p-5)}{6!}\Delta^6 y_0 + \ldots\right]$$

Now integrating it term by term, we get, after substituting the limits, as



$$\int_{x_0}^{x_0+nh} f(x)\,dx$$

$$= h\left[ny_0 + \frac{n^2}{2}\Delta y_0 + \frac{1}{2}\left\{\frac{n^3}{3} - \frac{n^2}{2}\right\}\Delta^2 y_0 + \frac{1}{6}\left\{\frac{n^4}{4} - n^3 + n^2\right\}\Delta^3 y_0\right.$$

$$+ \frac{1}{24}\left\{\frac{n^5}{5} - \frac{3n^4}{2} + \frac{11n^3}{3} - \frac{3n^2}{1}\right\}\Delta^4 y_0$$

$$+ \frac{1}{120}\left\{\frac{n^6}{6} - 2n^5 + \frac{35n^4}{4} - \frac{50n^3}{3} + 12n^2\right\}\Delta^5 y_0$$

$$\left.+ \frac{1}{720}\left\{\frac{n^7}{7} - \frac{15n^6}{6} + 17n^5 - \frac{225n^4}{4} + \frac{274n^3}{3} - 60n^2\right\}\Delta^6 y_0 + \ldots\right] \quad \ldots(1)$$

☛ **Q. 3. By means of Lagrange's formula prove that.**
$y_1 = y_3 - 0.3\,(y_5 - y_{-3}) + 0.2\,(y_{-3} - y_{-5})$

Ans. In this case y_1 is obtained while y_{-5}, y_{-3}, y_3, y_5 are given

$$y_1 = \frac{(1+3)(1-3)(1-5)}{(-5+3)(-5-3)(-5-5)}y_{-5} + \frac{(1+5)(1-3)(1-5)}{(-3+5)(-3-3)(-3-5)}y_{-3}$$

$$+ \frac{(1+5)(1+3)(1-5)}{(3+5)(3+3)(3-5)}y_3 + \frac{(1+5)(1+3)(1-3)}{(5+5)(5+3)(5-3)}y_5$$

$$y_1 = \frac{(-4)(-2)(-4)}{(-2)(-8)(-10)}y_{-5} + \frac{(6)(-2)(-4)}{(2)(-6)(8)}y_{-3}$$

$$+ \frac{(6)(4)(-4)}{(8)(6)(-2)}y_3 + \frac{(6)(4)(-2)}{(10)(8)(2)}y_5$$

$$y_1 = -\frac{1}{5}y_{-5} + \frac{1}{2}y_{-3} - \frac{3}{10}y_5 = \cdot 2y_{-5} + \cdot 5y_{-3} + y_3 - \cdot 3y_5$$

$$= y_2 - \cdot 3\,(y_5 - y_{-3}) + \cdot 2\,(y_{-3} - y_{-5}).$$

Q. 4. Derive the Weddle rule of numerical integration.

Ans.

Numerical Integration

The process of computing $\int_a^b y\,dx$, where $y = f(x)$ is given by a set of tabulated values $[x_i, y_i]$, $i = 0, 1, 2, \ldots, n$, $a = x_0$ and $b = x_n$ is called *numerical integration*.

Weddle's Rule

Taking $n = 6$ in Eqn (1) of § 3·2 above and neglecting all differences above the sixth order, we get

$$\int_{x_0}^{x_0+6h} f(x)\,dx = h\left[6y_0 + 18\Delta y_0 + 27\Delta^2 y_0 + 24\Delta^3 y_0 + \frac{123}{10}\Delta^4 y_0 + \frac{33}{10}\Delta^5 y_0 + \frac{41}{140}\Delta^6 y_0\right]$$

Replacing the coefficient of $\Delta^6 y_0$ by $\frac{42}{140}$ (the error made will be negligible) and the differences in terms of y's, we get

$$\int_{x_0}^{x_0+6h} f(x)\,dx = \frac{3h}{10}[y_0 + 5y_1 + y_2 + 6y_3 + y_4 + 5y_5 + y_6]$$

Similarly,

$$\int_{x_0+6h}^{x_0+12h} f(x)\,dx = \frac{3h}{10}[y_6 + 5y_7 + y_8 + 6y_9 + y_{10} + 5y_{11} + y_{12}]$$

$$\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots$$

$$\int_{x_0+(n-6)h}^{x_0+nh} = \frac{3h}{10}[y_{n-6} + 5y_{n-5} + y_{n-4} + 6y_{n-3} + y_{n-2} + 5y_{n-1} + y_n]$$

(if n is a multiple of 6)

Adding all these integrals and rearranging them in an order, we have

$$\int_{x_0}^{x_0+nh} f(x)\,dx = \frac{3h}{10}[(y_0 + y_n) + (5y_1 + y_2 + 6y_3 + y_4 + 5y_5) + (2y_6 + 5y_7 + y_8 + 6y_9 + y_{10} + 5y_{11}) + \ldots + (2y_{n-6} + 5y_{n-5} + y_{n-4} + 6y_{n-3} + y_{n-2} + 5y_{n-1})]$$

This is the Weddle's rule.

Q. 5. Explain Simpson's 3/8 rule for numerical integration.

Ans.

Simpson's three eighth's rule

Taking $n = 3$ in Eqn (1) and neglecting all differences above the third order, we get

$$\int_{x_0}^{x_0+3h} f(x)\,dx = h\left[3y_0 + \frac{9}{2}\Delta y_0 + \frac{1}{2}\left(\frac{27}{3} - \frac{9}{2}\right)\Delta^2 y_0\right.$$

$$+\frac{1}{6}\left(\frac{81}{4}-27+9\right)\Delta^3 y_0\Bigg]$$

$$= h\left[3y_0+\frac{9}{2}(y_1-y_0)+\frac{9}{4}(y_2-2y_1+y_0)+\frac{3}{8}(y_3-3y_2+3y_1-y_0)\right]$$

$$= \frac{3h}{8}[y_0+3y_1+3y_2+y_3]$$

Similarly, $\displaystyle\int_{x_0+3h}^{x_0+6h} f(x)\,dx = \frac{3h}{8}[y_3+3y_4+3y_5+y_6]$

... ...

$$\int_{x_0+(n-3)h}^{x_0+nh} f(x)\,dx = \frac{3h}{8}[y_{n-3}+3y_{n-2}+3y_{n-1}+y_n]$$

(if n is a multiple of 3)

Adding all these integrals, where n is a multiple of 3, we get

$$\int_{x_0}^{x_0+nh} f(x)\,dx = \frac{3h}{8}(y_0+y_n)+3(y_1+y_2+y_4+y_5+y_7+\ldots+y_{n-1})$$
$$+2(y_3+y_6+y_9+\ldots+y_{n-3})]$$

This is the Simpson's 3/8 rule. ●

☛ **Q. 6. Find the first derivative of $f(x)$ at $x = .4$ from the following table :**

x	.1	.2	.3	.4
$y = f(x)$	1.10517	1.22140	1.34986	1.49182

Ans. Since the derivative is found near the end so Newton's backward formula is used.

x	$f(x) = y$	Δy	$\Delta^2 y$	$\Delta^3 y$
·1	1·10517			
		·11623		
·2	1·22140		·01223	
		·12846		·00127
·3	1·34986		·01350	
		·14196		
·4	1·49182			

Newton's backward formula is

$$f(a+nh+xh) = f(a+nh)$$
$$+x\,\Delta f(a+nh)+\frac{x(x+1)}{2}\Delta^2 f(a+nh)$$
$$+\frac{x(x+1)(x+2)}{6}\Delta^3 f(a+nh)+\ldots \quad \ldots(1)$$

Here $f(a+nh)$ is the last term $h = \cdot 1, a = \cdot 1, n = 3$

$\Rightarrow \quad x = 0$

Differentiating (1) w.r.t. x, we get

$$hf'(a+nh+xh) = \Delta f(a+nh) + \frac{2x+1}{2}\Delta^2 f(a+nh) + \frac{3x^2+6x+2}{6}\Delta^3 f(a+nh) + \ldots \quad \ldots(2)$$

On substituting $x = 0$, we get

$$(\cdot 1) f'(\cdot 4) = \Delta f(\cdot 4) + \frac{1}{2}\Delta^2 f(\cdot 4) + \frac{1}{3}\Delta^3 f(\cdot 4)$$

$$(\cdot 1) f'(\cdot 4) = \cdot 14196 + \frac{1}{2}(\cdot 01350) + \frac{1}{3}(\cdot 00126) = \cdot 14913$$

or $\quad f'(\cdot 4) = 1 \cdot 4913.$

☛ **Q. 7. The velocity v of a particle at distance s from a point on its path is given by the following table :**

s (ft)	0	10	20	30	40	50	60
v [ft/s]	47	58	64	65	61	52	38

Estimate the time taken to travel 60 ft using Simpson's 1/3 rule. Compare the result with Simpson's 3/8 rule.

Ans. We know that the rate of displacement is velocity, *i.e.*, $ds/dt = v$, Therefore, the time taken to travel 60 ft is given by

$$t = \int_0^{60} \frac{1}{v}\, ds = \int_0^{60} y\, dx$$

where $s = x$ and $y = 1/v$. The table is as given below :

$x\,(=s)$	0	10	20	30
$y = 1/v$	0.0212765 $= y_0$	0.0172413 $= y_1$	0.015625 $= y_2$	0.0153846 $= y_3$
$x\,(=s)$	40	50	60	
$y = 1/v$	0.0163934 $= y_4$	0.0192307 $= y_5$	0.0263157 $= y_6$	

By Simpson's 1/3 rule :

$$\int_0^{60} y\, dx = \frac{h}{3}[(y_0+y_6) + 2(y_2+y_4) + 4(y_1+y_3+y_5)]$$

$$= \frac{10}{3}[(0.0212765 + 0.0263157) + 2(0.015625 + 0.0163934) + 4(0.0172413 + 0.0153846 + 0.0192307)]$$

$$= 1.063518$$

Hence, the time taken to travel 60 ft is 1.064s.

By Simpson's 3/8 rule :

$$\int_0^{60} y\,dx = \frac{3h}{8}[(y_0 + y_6) + 3(y_1 + y_2 + y_4 + y_5) + 2y_3]$$

$$= \frac{30}{8}[(0.0212765 + 0.0263157) + (0.0172413 + 0.015625 + 0.0163934 + 0.0192307) + 2(0.0153846)]$$

$$= 1.0643723$$

By this method also the time taken to travel 60 ft is 1.064s.

Q. 8. Find the root of the equation $2x = \cos x + 3$ correct to three decimal places.

Ans. $x = \frac{1}{2}(\cos x + 3) \therefore \phi(x) = \frac{1}{2}(\cos x + 3)$

$$\phi'(x) = \frac{1}{2}(-\sin x)$$

$$\therefore |\phi'(x)| = \frac{1}{2}|-\sin x|$$

$\therefore$ $|\phi'(x)| < 1$ max value of $\sin x$ is 1

So iterative method is applicable and starting with $x_0 = \pi/2$

$$x_1 = \frac{1}{2}.3 = \frac{3}{2} = 1 \cdot 5$$

$$x_2 = \frac{1}{2}(\cos 1 \cdot 5 + 3) = \frac{1}{2}(0 \cdot 0707 + 3) = 1 \cdot 535$$

$$x_3 = \frac{1}{2}(\cos 1 \cdot 535 + 3) = \frac{1}{2}(0 \cdot 0354 + 3) = 1 \cdot 5177$$

$$x_4 = \frac{1}{2}(\cos 1 \cdot 5177 + 3) = \frac{1}{2}(0 \cdot 05306 + 3) = 1 \cdot 5265$$

$$x_5 = \frac{1}{2}(\cos 1 \cdot 5265 + 3) = \frac{1}{2}(0 \cdot 0442 + 3) = 1 \cdot 5221$$

$$x_6 = \frac{1}{2}(\cos 1 \cdot 5221 + 3) = \frac{1}{2}(0 \cdot 0486 + 3) = 1 \cdot 5243$$

$$x_7 = \frac{1}{2}(\cos 1 \cdot 5243 + 3) = \frac{1}{2}(0 \cdot 0464 + 3) = 1 \cdot 5232$$

$$x_8 = \frac{1}{2}(\cos 1 \cdot 5232 + 3) = \frac{1}{2}(0 \cdot 0475 + 3) = 1 \cdot 5232$$

Hence the required root is $1 \cdot 5232$.

Q. 9. Find the cube root of 10 by itertive method.

Ans. The iterative formula for finding successive approximation of cube roots is given by

$$x_{n+1} = \frac{1}{3}\left(2x_n + \frac{N}{x_n^2}\right)$$

$\because$ $(8)^{1/3} = 2$ and $(27)^{1/3} = 3$ 10 lies within (8, 27)

$\therefore$ We may take $x_0 = \dfrac{2+3}{2} = 2\cdot5$

$$x_1 = \frac{1}{3}\left(2 \times 2\cdot5 + \frac{10}{(2\cdot5)^2}\right) = \frac{1}{3}\left(5 + \frac{10}{6\cdot25}\right) = 2\cdot2$$

$$x_2 = \frac{1}{3}\left(2 \times 2\cdot2 + \frac{10}{4\cdot84}\right) = 2\cdot155$$

and $$x_3 = \frac{1}{3}\left(2 \times 2\cdot155 + \frac{10}{(2\cdot155)^2}\right) = 2\cdot15466$$

$x_2 \simeq x_3$. Hence the cube root of $2\cdot15466$

$\therefore \sqrt[3]{10} \simeq 2\cdot15466$

$\therefore \sqrt[3]{10} \simeq 2\cdot15466.$

Q. 10. Find a real root of the equation

$$f(x) = x^3 + x^2 - 1 = 0$$

by using iteration method.

Ans. Here $f(0) = -1$ and $f(1) = 1 + 1 - 1 = 1$ so that the root lies between 0 and 1

$x^3 + x^2 - 1 = 0$ $x^2(x+1) = 1$

$$\therefore \quad x = \frac{1}{\sqrt{x+1}} = \phi(x)$$

$$\therefore \quad \phi'(x) = \frac{-1}{2(x+1)^{3/2}}$$

$\therefore$ $|\phi'(x)| < 1$ for $x < 1$

So taking $x_0 = 0\cdot5$ average of 0 and 1.

Hence iterative method is applicable

$$x_1 = \phi(x) = \frac{1}{\sqrt{x+1}} = \frac{1}{\sqrt{0\cdot5}} = \cdot81649$$

$$x_2 = \frac{1}{\sqrt{1\cdot81649}} = \cdot74196$$

$$x_3 = \frac{1}{\sqrt{1\cdot74196}} = \cdot75767$$

$$x_4 = \frac{1}{\sqrt{1 \cdot 75767}} = \cdot 75427$$

$$x_5 = \frac{1}{\sqrt{1 \cdot 75427}} = \cdot 75500$$

$$x_6 = \frac{1}{\sqrt{1 \cdot 75500}} = \cdot 75485$$

$$x_7 = \frac{1}{\sqrt{1 \cdot 75485}} = \cdot 75488$$

$$x_8 = \frac{1}{\sqrt{1 \cdot 75488}} = \cdot 75487$$

as $x_7 \simeq x_8$. So the required root is $\cdot 75487$. ●

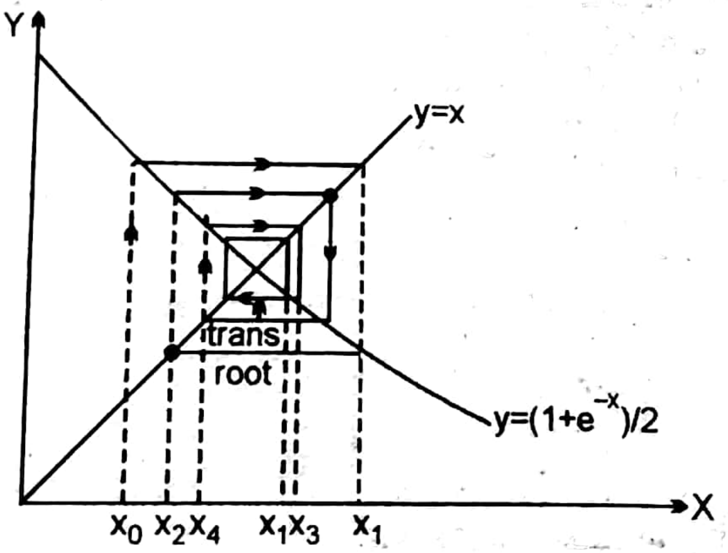
Q.11. Discuss Method of Successive Approximation or Iteration.

Ans. Method of Successive Approximation or Iteration

If the equation $f(x) = 0$ whose roots are to be found is expressed as $x = g(x)$ then an iterative method is as follows

e.g. $$f(x) = e^{-x} = 2x - 1 = 0$$

so let $x = (1 + e^{-x})/2$



choose any guess $x = x_0$ putting in $(1 + e^{-x})/2$, then $x_1 = (1 + e^{-x_0})/2$ again putting x_1 in $(1 + e^{-x})/2$ and so on until the value converges

The successive approximation method is first order convergent as shown below

$$x_{i+1} = g(x_i)$$

Let y be the root, e_i the error in the i^{th} iteration and e_{i+1} the error in $i + 1^{th}$ iteration then

$$x - e_i = y_i \quad \text{or} \quad x_i - e_i = y$$

$$y + e_{i+1} = g(y + e_i)$$

$$y + e_{i+1} = g(y) + e_i \left.\frac{dg}{dx}\right|_y + \text{higher orders}$$

or $$e_{i+1} = e_i \left.\frac{dg}{dx}\right|_y + (g(y) - y) \text{ as we know } y = g(y)$$

$$\lim_{i \to \infty} \left|\frac{e_{i+1}}{e_i}\right| = \text{constt} = \left.\frac{dg}{dx}\right|_y$$

●

☛ **Q. 12. Let $x = \alpha$ be a root of $f(x) = 0$, which is equivalent to $x = \phi(x)$ and I be an interval containing α. If $|\phi'(x)| < 1$, for all x in I, then the sequence of approximations $x_0, x_1, \ldots, x_n$ will converge to the root α, provided that the initial approximation x_0 is chosen in I.**

Ans. α is a root of $f(x) = 0$

$\Rightarrow$ α is a root of $x = \phi(x)$

$\Rightarrow$ $\alpha = \phi(\alpha)$. ...(3)

Let x_{n-1} and x_n denote two successive approximations to α, then we have

$$x_n = \phi(x_{n-1}),$$

$$\therefore\ x_n - \alpha = \phi(x_{n-1}) - \phi(\alpha).$$

By mean value theorem we have

$$\frac{\phi(x_{n-1}) - \phi(\alpha)}{x_{n-1} - \alpha} = \phi'(\beta),$$

where $x_{n-1} < \beta < \alpha$

$$\Rightarrow \phi(x_{n-1}) - \phi(\alpha) = (x_{n-1} - \alpha)\phi'(\beta). \quad \ldots(4)$$

Let λ be the maximum absolute value of $\phi'(x)$ over I, then from (4), we have

$$|x_n - \alpha| \le \lambda |x_{n-1} - \alpha| \quad \ldots(5)$$

$$\Rightarrow |x_{n-1} - \alpha| \le \lambda |x_{n-2} - \alpha|$$

$$\therefore\ |x_n - \alpha| \le \lambda^2 |x_{n-2} - \alpha| \quad \ldots(6)$$

Proceeding in this way, we get

$$|x_n - \alpha| \le \lambda^n |x_0 - \alpha| \quad \ldots(7)$$

If $\lambda < 1$ over I, then the RHS of (7) becomes small (as n increases) such that

$$\lim |x_n - \alpha| = 0$$

or $$\lim x_n = \alpha.$$

$\therefore$ The sequence of approximations will converge to the root α if $\lambda < 1$ i.e., $|\phi'(x)| < 1$.

If $\lambda > 1$, then $|\phi'(x)| > 1$

$\Rightarrow$ $|x_n - \alpha|$ will become indefinitely large, as n increases and the sequence of approximations does not converge.

●

Q. 13. Find a real root of cosx = 3x – 1, correct to three decimal places.

Ans. We have

$$f(x) = \cos x - 3x + 1 = 0$$

$$f(0) = \cos 0 - 0 + 1 = 2 > 0$$

$$f\left(\frac{\pi}{2}\right) = \cos\frac{\pi}{2} - 3\frac{\pi}{2} + 1 = -\frac{3\pi}{2} + 1 < 0$$

$$\Rightarrow f(0) \cdot f\left(\frac{\pi}{2}\right) < 0$$

$\therefore$ A root lies between 0 and $\frac{\pi}{2}$

The given equation can be written as $x = \frac{1}{3}[1 + \cos x]$

Here $\phi(x) = \frac{1}{3}[1 + \cos x]$

$$\therefore \quad \phi'(x) = \frac{-\sin x}{3}$$

$$\Rightarrow \quad |\phi'(x)| = \left|\frac{\sin x}{3}\right| < 1 \text{ in } \left(0, \frac{\pi}{2}\right)$$

$\therefore$ Iteration method can be applied.

Let $x_0 = 0$

be the initial approximation.

$\therefore$ We get

$$x_1 = \phi(x_0) = \frac{1}{3}[1 + \cos 0] = 0.66667$$

$$x_2 = \phi(x_1) = \frac{1}{3}[1 + \cos 0.66667] = 0.595295 \approx 0.59530$$

$$x_3 = \phi(x_2) = \frac{1}{3}[1 + \cos 0.59530] = 0.6093267 \approx 0.60933$$

$$x_4 = \phi(x_3) = \frac{1}{3}[1 + \cos 0.60933] = 0.6066771 \approx 0.60668$$

$$x_5 = \phi(x_4) = \frac{1}{3}[1 + \cos 0.60668] = 0.6071818 \approx 0.60718$$

$$x_6 = \phi(x_5) = \frac{1}{3}[1 + \cos 0.60718] = 0.6070867 \approx 0.60709$$

$$x_7 = \phi(x_6) = \frac{1}{3}[1 + \cos 0.60709] = 0.6071039 \approx 0.60710$$

$$x_8 = \phi(x_7) = \frac{1}{3}[1 + \cos 0.60710] = 0.60710$$

The correct root of the equation is 0.607 correct to three decimal places.

Q. 14. Find by the iteration method, the root near 3.8, of the equation $2x - \log_{10}x = 7$ correct to four decimal places.

Ans. The given equation can be written as $x = \frac{1}{2}[\log_{10}x + 7]$

Clearly $|\phi'(x)| < 1$ when x is near 3.8

We have $x_0 = 3.8$

$$x_1 = \phi(x_0) = \frac{1}{2}[\log_{10} 3.8 + 7] = 3.79,$$

$$x_2 = \phi(x_1) = \frac{1}{2}[\log_{10} 3.79 + 7] = 3.7893,$$

$$x_3 = \phi(x_2) = \frac{1}{2}[\log_{10} 3.7893 + 7] = 3.7893,$$

$$x_2 = x_3 = 3.7893$$

∴ We can take 3.7893 as the root of the given equation.

●

Q. 15. Discuss the Newton-Raphson Method.

Ans. Newton-Raphson method : Let x_0 be an approximate root of the equation $f(x) = 0$. If $x_1 = x_0 + h$ be the exact root, then $f(x_1) = 0$.

∴ Expanding $f(x_0 + h)$ by Taylor's series

$$f(x_0) + hf'(x_0) + \frac{h^2}{2!}f''(x_0) + \ldots\ldots\ldots\ldots = 0$$

Since h is small, neglecting h^2 and higher powers of h, we get

$$f(x_0) + hf'(x_0) = 0$$

or

$$h = -\frac{f(x_0)}{f'(x_0)} \qquad ..(1)$$

∴ A closer approximation to the root given by

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Similarly, starting with x_1, a still better approximation x_2 is given by

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$

In general,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad ..(2)$$

which is known as the Newton-Raphson formula or Newton's iteration formula.

●

Q. 16. Find by Newton's method the real root of the equation $3x = \cos x + 1$

Ans. Let $f(x) = 3x - \cos x - 1$

$f(0) = -2 =$ ve,

$f(1) = 3 - 0.5403 - 1 = 1.4597 = +$ ve.

So a root of $f(x) = 0$ lies between 0 and 1. It is nearer to 1. Let us take $x_0 = 0.6$.

Also, $f'(x) = 3 + \sin x$

∴ Newton's iteration formula gives

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{3x_n - \cos x_n - 1}{3 + \sin x_n}$$

$$= \frac{x_n \sin x_n + \cos x_n + 1}{3 + \sin x_n} \qquad ..(1)$$

Putting $n = 0$, the first approximation x_1 is given by

$$x_1 = \frac{x_0 \sin x_0 + \cos x_0 + 1}{3 + \sin x_0} = \frac{(0.6)\sin(0.6) + \cos(0.6) + 1}{3 \sin(0.6)}$$

$$= \frac{0.6 \times 0.5729 + 0.82533 + 1}{3 + 0.5729} = 0.6071$$

Putting $n = 1$ in (1), the second approximation is

$$x_2 = \frac{x_1 \sin x_1 + \cos x_1 + 1}{3 + \sin x_1}$$

$$= \frac{0.6071 \sin(0.6071) + \cos(0.6071) + 1}{3 + \sin(0.6071)}$$

$$= \frac{0.6071 \times 0.57049 + 0.8213 + 1}{3 + 0.57049} = 0.6071$$

Clearly $x_1 = x_2$

Hence the desired root is 0.6071 correct to four decimal places.

●

☛ Q. 17. Using Newton's iterative method, find the real root of $x \log_{10} x = 1{\cdot}2$ correct to five decimal places.

Ans. Let $f(x) = x \log_{10} x - 1.2$; $f(1) = -1.2 = -$ ve,

$f(2) = 2 \log_{10} 2 - 1.2 = 0.59794 = -$ ve

and $f(3) = 3 \log_{10} 3 - 1.2 = 1.4314 - 1.2 = 0.23136 = +$ ve

So a root of $f(x) = 0$ lies between 2 and 3. Let us take $x_0 = 2$.

Also $f'(x) = \log_{10} x + x \cdot \frac{1}{x} \log_{10} e = \log_{10} x + 0.43429.$

∴ Newton's interation formula gives

$$\{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = \frac{0.43429\, x_n + 1.2}{\log_{10} x_n + 0.43429} \qquad ..(1)$$

Putting $n = 0$, the first approximation is

$$x_1 = \frac{0.43429 \times x_0 + 1.2}{\log_{10} x_0 + 0.43429} = \frac{0.43429 \times 2 + 1.2}{\log_{10} 2 + 0.43429} = \frac{0.86858 + 1.2}{0.30103 + 0.43429} = 2.81$$

Similarly putting $n = 1, 2, 3, 4$ in (1), we get

$$x_2 = \frac{0.43429 \times 2.81 + 1.2}{\log_{10} 2.81 + 0.43429} = 2.741 \quad x_3 = \frac{0.43429 \times 2.741 + 1.2}{\log_{10} 2.81 + 0.43429} = 2.74064$$

$$x_4 = \frac{0.43429 \times 2.74064 + 1.2}{\log_{10} 2.74064 + 0.43429} = 2.74065$$

$$x_5 = \frac{0.43429 \times 2.74065 + 1.2}{\log_{10} 2.74065 + 0.43429} = 2.74065$$

Clearly $x_4 = x_5$.

Hence the required root is 2. 74065 correct to five decimal places.

Q. 18. Calculate the total arc of the ellipse $4x^2 + y^2 = 4$.

Ans. We know that total arc length of the the ellipse $= s$

$$= 4\int_0^1 \sqrt{\{[1 + (dy/dx)^2]\}}\, dx = u\int_0^1 \sqrt{\{[1 + (-4x/y)^2]\}}\, dx$$

(as $dy/dx = -4x/y$)

$$= 4\int_0^1 \sqrt{[(1 + 3x^2)/(1 - x^2)]}\, dx.$$

Putting $x = \sin\theta$, we get

$$s = 4\int_0^{\pi/2} \sqrt{[1 + 3\sin^2\theta]}\, d\theta = 4\int_0^{\pi/2} f(\theta)\, d\theta.$$

Now divide the interval $[0, \pi/2]$ in 6 equal parts by points $0, \pi/12, \pi/6, \pi/4, \pi/3, 5\pi/12, \pi/2$. Then

$$f(0) = 1, f(\pi/12) = \sqrt{\{[1 + 3\sin^2(\pi/12)]\}} = \sqrt{[1 + 3(0\cdot2588)^2]}$$
$$= \sqrt{[1\cdot2009]} = 1\cdot095$$

Similarly, $f(\pi/3) = 1\cdot323, f(\pi/4) = 1\cdot581, f(\pi/3) = 1\cdot803,$
$f(5\pi/12) = 1\cdot949, f(\pi/2) = 2.$

We are given 7 ordinates, so we can use Weddle's rule. We obtain

$$\int_0^{\pi/2} \sqrt{[1 + 3\sin^2\theta]}\, d\theta = (3/10)\,.\,(\pi/12)\,[(1 + 2)$$
$$+ 5(1\cdot095 + 1\cdot949) + (1\cdot323 + 1\cdot803) + 6 \times 1\cdot581]$$
$$= (\pi/40)\,[3 + 15\cdot22 + 3\cdot126 + 9\cdot486] = 2\cdot422.$$

$$\Rightarrow \quad s = 4 \times 2\cdot422 = 9\cdot688.$$

Q. 19. Compute the value of the definite integral $\int_{0\cdot2}^{1\cdot4} (\sin x - \log_e x + e^x)\, dx$ by (i) Trapezoidal rule (ii) Simpson's 1/3–rule (iii) Simpson's 3/8–rule (iv) Weddle rule. After finding the true value of the integral, compare the error in the four cases.

Ans. Let $y = \sin x - \log_e x + e^x$. Then taking $h = \cdot2$, the values of y are computed for different values of x. The values of integral are as follows :

x_4	$\sin x_i$	$\log_e x_i$	e^{ξ}	y_i	Calculations for			
					Trapezium rule	Simpson's 1/3 rule	Simpson's 3/8 rule	Weddle rule
$x_o = \cdot 2$	0·1987	– 1·6095	1·2214	3·0296	$y_0 = 3\cdot 0296$	$y_0 = 3\cdot 0296$	$y_0 = 3\cdot 0296$	$y_0 = 3\cdot 0296$
$x_1 = \cdot 4$	0·3894	– 0·9163	1·4918	2·7975	$2y_1 = 5\cdot 5950$	$4y_1 = 11\cdot 1800$	$3y_1 = 8\cdot 3925$	$5y_1 = 13\cdot 9875$
$x_2 = \cdot 6$	0·5646	– 0·5108	1·8221	2·8975	$2y_2 = 5\cdot 7950$	$2y_2 = 5\cdot 7950$	$3y_2 = 8\cdot 6925$	$y_2 = 2\cdot 8975$
$x_3 = \cdot 8$	0·7174	– 0·2232	2·2255	3·1661	$2y_3 = 6\cdot 3322$	$4y_3 = 12\cdot 6644$	$2y_3 = 6\cdot 3322$	$6y_3 = 18\cdot 9966$
$x_4 = 1\cdot 0$	0·8415	0·0000	2·7183	3·5598	$2y_4 = 7\cdot 1196$	$2y_4 = 7\cdot 1196$	$3y_4 = 10\cdot 6794$	$y_4 = 3\cdot 5598$
$x_5 = 1\cdot 2$	0·9320	0·1823	3·3201	4·0698	$2y_5 = 8\cdot 1396$	$4y_5 = 16\cdot 2792$	$3y_5 = 12\cdot 2094$	$5y_5 = 10\cdot 3490$
$x_6 = 1\cdot 4$	0·9855	0·3365	4·0552	4·7042	$y_6 = 4\cdot 7042$	$y_6 = 4\cdot 7042$	$y_6 = 4\cdot 7042$	$y_6 = 4\cdot 7042$
					40·7152	60·7820	54·0398	67·5242

(i) By Trapezoidal rule, $I = (1/2)\ (0\cdot 2) \times 40\cdot 7152 = 4\cdot 0715$.

(ii) By Simpson's '1/3'–rule,

$$I = (1/3)\ (0\cdot 2) \times 60\cdot 7820 = 4\cdot 0521.$$

(iii) By Simpson's '3/8'–rule,

$$I = (3/8)\ (0\cdot 2) \times 54\cdot 0398 = 4\cdot 0529$$

(iv) By Weddle's rule, $I = (3/10)\ (0\cdot 2) \times 67\cdot 5242 = 4\cdot 0515$.

The actual value of the intergral $= I_T$

$$= \Big[-\cos x - (x \log_e x - x) + e^x\Big]_{0\cdot 2}^{1\cdot 4}$$

$$= -[\cos 1\cdot 4 - \cos 0\cdot 2] - [1\cdot 4 \log_e 1\cdot 4 - 1\cdot 4 - 0\cdot 2 \log 0\cdot 2 + 0\cdot 2] + (e^{1\cdot 4} - e^{0\cdot 2})$$

$$= -(0\cdot 1700 - 0\cdot 9801) - [(1\cdot 4 \times 0\cdot 3365 - 1\cdot 2 - (0\cdot 2)\ (-1\cdot 6095)] + (4\cdot 0552 - 1\cdot 2214)$$

$$= 0\cdot 801 + 0\cdot 4070 + 2\cdot 8338 = 4\cdot 0509$$

$\therefore$ Error term in case (i) $= 4\cdot 0715 - 4\cdot 0509 = 0\cdot 0206$,

Error term in case (ii) $= 4\cdot 0521 - 4\cdot 0509 = 0\cdot 0012$

Error term in case (iii) $= 4\cdot 0529 - 4\cdot 0509 = 0\cdot 0020$

Error term in case (iv) $= 4\cdot 0515 - 40509 = 0\cdot 0006$.

Q. 20. From the following values of $y = f(x)$ in the given range of values of x, find the position of the centroid of the area under the curve and the x-axis.

x	0	1/4	1/2	3/4	1
y	1	4	8	4	1

Ans. The centroid of the plane area under the curve $y = f(x)$ is given by $(\bar{x}, \bar{y})$, where

$$\bar{x} = \int_0^1 xy\, dx \div \int_0^1 y\, dx;$$

$$\bar{y} = \int_0^1 (y^2/2)\, dx \div \int_0^1 y\, dx \qquad \text{...(i)}$$

From the given data, we obtain

xy :	0	1	4	3	1
$y^2/2$:	1	8	32	8	1

$\therefore$ By Simpson's '1/3' rule

$$\int_0^1 xy\, dx = (1/3)\ (1/4)\ [(0 + 1) + 4 \times (1 + 3) + 2 \times 4]$$
$$= 25/12$$

$$\int_0^1 (y^2/2)\, dx = (1/3)\ (1/4)\ [(1 + 1) + 4 \times (8 + 8) + 2 \times 32]$$
$$= 130/12.$$

$$\int_0^1 y\,dx = (1/3)\,.\,(1/4)\,[(1+1) + 4\times(4+4) + 2\times 8]$$
$$= 50/12$$

$\therefore$ From (i),

$$\bar{x} = \frac{25}{12} \div \frac{50}{12} = \frac{1}{2};\ \bar{y} = \frac{130}{12} \div \frac{50}{12} = \frac{13}{5}$$

$\Rightarrow$ the centroid is the point (1/2, 13/5) *i.e.,* $(0\cdot5, 2\cdot6)$.

Ex. 21. Prove Simpson's formula

$$\int_a^b f(x)\,dx = \frac{b-a}{6n}\,[f(x_0) + 4f(x_1) + 2f(x_2) + \ldots + f(x_{2n})]$$

where $x_0 = a, x_{2n} = b$

and use it to evaluate $\int_1^2 \frac{dx}{x}$ and give estimates of error for $n = 1$ and 2, given that $\log_e 2 = 0\cdot69315$.

Ans. Simpson's formula is

$$\int_{x_0}^{x_0+nh} f(x)\,dx = \frac{h}{3}\,[f(x_0) + f(x_0+nh) + 4\,\{f(x_0+h) + f(x_0+3h) + \ldots\} + 2\,\{f(x_0+2h) + f(x_0+4h) + \ldots\}]$$

But here $x_0 = a$ and $x_0 + 2nh = x_{2n} = b$.

Also $h = (b-a)/2n$. Putting all these values, we get the desired form of Simpson's rule.

Now when $n = 1$, it means we should divide the whole range (1, 2) in two equal parts by three ordinates x_0, x_1, x_2. Then by the above formula

$$\int_0^2 \frac{dx}{x} = \int_{x_0}^{x_2} f(x)\,dx \qquad \text{where } f(x) = \frac{1}{x},\ x_2 = 2,\ x_0 = 1$$
$$= \frac{b-a}{6n}\,[f(x_0) + 4f(x_1) + f(x_2)]$$
$$= \frac{2-1}{6\times 1}\left[\frac{1}{1} + \frac{4\cdot1}{3/2} + \frac{1}{2}\right] = \frac{1}{6}\cdot\frac{25}{6} = \cdot69444.$$

Similarly putting $n = 2$ and dividing the range (1, 2) in four equal parts by five ordinates $x_0 = 1, x_1, x_2, x_3, x_4 = 2$, we get

$$\int_1^2 \frac{dx}{x} = \int_{x_0}^{x_4} f(x)\,dx = \frac{b-a}{6n}\,[f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + f(x_4)]$$

$$= \frac{1}{12}\left[\frac{1}{1} + 4 \cdot \frac{1}{5/4} + 2 \cdot \frac{1}{3/2} + 4 \cdot \frac{1}{7/4} + \frac{1}{2}\right] = \cdot 69321.$$

Now the actual value is

$$\int_1^2 \frac{dx}{x} = \Big[\log_e x\Big]_1^2 = \log_e 2 = \cdot 69315.$$

Hence the error when $n = 1$, is $\cdot 69315 - \cdot 69444 = - \cdot 00129$ and when $n = 2$, is $\cdot 69315 - \cdot 69321 = - \cdot 00006$.

●

B.A./B.Sc. (Part–I) STATISTICS, 2015
(Second Paper)
Probability Distribution & Numerical Analysis

Note : Answer five questions in all selecting one question from each Unit and Question No. 1, which is compulsory.

1. (a) The distribution of a variable x is given by the following law :

$f(x) = \text{constant}.\, e^{-\frac{1}{2}\left(\frac{x-100}{5}\right)^2}, -\infty < x < \infty$. Find the value of

(i) Constant

(ii) Mean and

(iii) Variance

(b) Under what condition Binomial distribution tends to Normal distribution?

(c) Define negative binomial distribtion.

(d) Show that t-distribution becomes Cauchy distribution for $= 1$.

(e) What is χ^2-variant? Write down its probability density function.

(f) Define F-statistic and give its p.d.f.

(g) Define interpolation and write down the fundamental assumptions for interpolation.

(h) Explain Δ and E operators and prove that $E \equiv 1 + \Delta$.

(i) State fundamental assumptions of numerical analysis procedures.

(j) Find the value of $\log_e 7$ using Simpson's 1/3 rule.

UNIT–I

2. (a) Discuss main features of a normal distribution. Show that for normal distribution the mean deviation about mean is $\sigma\sqrt{\frac{2}{\pi}}$, where σ^2 is the variance of the distribution.

(b) Derive Poisson distribution as a limiting case of binomial distribution. Find its m.g.f., mean and variances.

3. (a) Define Exponential distribution and obtain its mean and variance.

(b) For the rectangular distribution :

$$f(x) = \frac{1}{2a}; -a \le x \le a$$

Show that $\mu_{2n} = \dfrac{a^{2n}}{(2n+1)}$.

UNIT–II

4. (a) Show that square of t-variable with n degrees of freedom is distribution as F with I and n degrees of freedom.

(b) Define Bivariat normal distribution. Find the marginal and conditional p.d.f. of this distribution.

5. (a) Define t-distribution and show that as $n \to \infty$, t distribution tends to normal distribution.

(b) State and prove the additive property of χ^2-distribution.

UNIT–III

6. State and prove Lagrange's interpolation formula. Use this formula to prove that :

$$y_3 = 0.05\,(y_0 + y_6) - 0.3\,(y_1 + y_5) + 0.75\,(y_2 + y_4)$$

7. (a) State and prove Newton's forward formula for interpolation.

(b) State and prove Gauss's central difference formula.

UNIT–IV

8. (a) Describe the method of numerical integration and obtain the general quadrature formula.

(b) Describe the Trapezoidal rule for numerical integration.

9. (a) Explain Simpson's 3/8 rule for numerical integration.

(b) Derive Weddel's rule of numerical integration.

●